

Transcription Start Sites Project Report

Sample TSS report for *onecut*

Student name(s): Wilson Leung
Student email(s): student@example.edu
Faculty advisor(s): Sarah C. R. Elgin
College/university: Washington University in St. Louis

Project details

Project name: contig35
Project species: *D. biarmipes*
Date of submission: 12/26/2023
Number of genes in project: 3

Does this report cover TSS annotations for all of the genes or is it a partial report? Partial report
If this is a partial report, please indicate the region of the project covered by this report:

From base 12,051 to base 21,700

Transcription start sites (TSS) report form

Gene name (e.g., *D. biarmipes eyeless*): *D. biarmipes onecut*

Gene symbol (e.g., *dbia_ey*): *dbia onecut*

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
<i>onecut-RA</i>	<i>onecut-RB</i>

Complete this report form for each gene in your project. Copy and paste this form to create as many copies as needed.

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:

NA

Isoform TSS report

Complete an Isoform TSS report (through page 7) for each unique TSS listed in the table above. If the gene has more than one unique TSS, copy and paste this form to create as many copies as needed.

Gene-isoform name (e.g., *dbia_ey-RA*): *dbia onecut-RA*

Names of the isoforms with the same TSS as this isoform:

dbia onecut-RB

Type of core promoter in *D. melanogaster* (see table below):

(Peaked / Intermediate / Broad / Insufficient Evidence)

Peaked

The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

Type of core promoter	# annotated TSS	# DHS positions
Peaked	1	0
	0	1
	1	1
Intermediate	≤ 1	> 1
	> 1	≤ 1
Broad	> 1	> 1
Insufficient Evidence	0	0

1. Annotate the first transcribed exon

Coordinates of the first transcribed exon based on *blastn* alignment:

21,599–18,924

Does the *blastn* alignment cover the entire *D. melanogaster* first transcribed exon? If not, specify the parts of the *D. melanogaster* exon that are missing from the *blastn* alignment.

Yes, the *blastn* alignment covers the entire length (2647 bp) of the first transcribed exon oncut:3

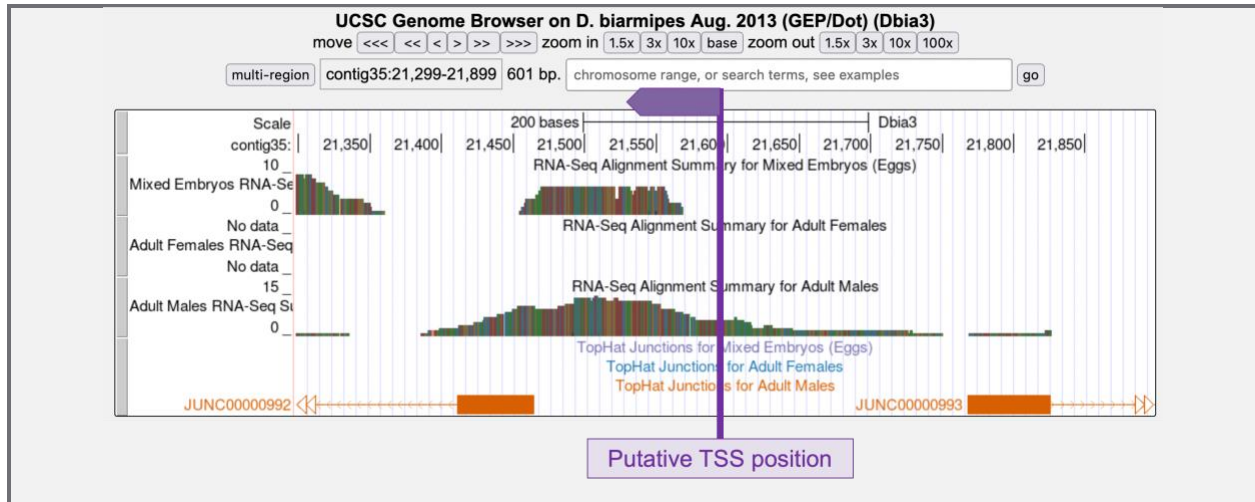
If the TSS annotation is supported by *blastn* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *blastn* alignment into the box below:**

contig35									
Sequence ID: Query_25615 Length: 48011 Number of Matches: 8									
Range 1: 18924 to 21599 Graphics					▼ Next Match ▲ Previous Match				
Score	Expect	Identities	Gaps	Strand	Score	Expect	Identities	Gaps	Strand
1509 bits(1055)	0.0	1953/2749(71%)	175/2749(6%)	Plus/Minus					
Query 1	CAGTTT	CGAATTTGGGTATAATTCGCGTGAGGATAAGTCTTCATGGAACGGCGGTCACAT	60	Query 1294	ATAGGCAGCATATGCAGCATGGCGGTGAG-----GTCAATACCAGTCCAGTGGACTTTG	1347			
Sbjct 21599	CAGTTT	CGAATTTAAGTATAATTCGCGCGAGGCTGAGACTTATGTAAGCAACGGCAGTCAAT	21540	Sbjct 20262	ATAGTCAAAACACGCGCAATGCCGACGCGCATAGTGTAGCAGCCAGTGGACTTTG	20203			
Query 61	AACAACATTT	CGGAAAAACACCCACATGTG---TGTACATGAGTGATAACCGTGCAATG	117	Query 1348	TATCATCAGATATAAATTTGGATGGTCTGACTGTAGACGACAGCGTCTCCAGACGGATC	1407			
Sbjct 21539	AACAACATTT	CGGAAAAACACCCACATGTGAAGTGTACATATGTGTATACACGAAGAG	21480	Sbjct 20202	TAGCTCCGACATAGCTTGGATGGTCTGACAGTAGATTCT---GATGTC---CAATCGGTC	20146			
Query 118	CGTAGAAAAATTT	CTAGCGGAAGTTTCAGTTGTGAAAAACAGACGAGGCACTGTGCGGGA	177	Query 1408	ACTCTCAAGAAACGGGTAAAGCAGGAGCAAACTACTTATTGTGCAATCAAGAGCT	1467			
Sbjct 21479	CGAAGATAAATTT	CTAGCGGAAGATTTCAGTTGTGACATGAACAGGCGTTGCGGCGAGA	21420	Sbjct 20145	TCTCCAGAGAGATGGCGATAAAGCAGGAGCAGAGGCTACTTATTGCCCAATCGAAGGCG	20086			
Query 178	GGCTCTAGGTTT	TTCATCAGTTTGATCCAGTTAAAGGTACCTTTTTTTAAAAATCAAC	237	Query 1468	AGGATCAAAAGCATAGGCGGATTCGAAATGCTTTGAGAGCTTTGAGCGTAAATTCGGCC	1527			
Sbjct 21419	GCCCATAGGTT	CTATCAGTTTCAACCAGTCAAAAGCTGCCCTCTTTTGGAAAAAG-ACAT	21361	Sbjct 20085	TGGATCAAGGCCATTAAGCGCATTCGAAATGCAGCTGGAAATGGCGGAATGTAATGCAAGTC	20026			
Query 238	CCTCTTAGAGCAAGT	CCCAATACATTAATAGTCTTCGTCTGATTGCAATGAAAACTTG	297	Query 1528	TTGGCTGCAGCTGGATGATGGATGAGATTCTCTCTGATGGCTGGCTGGCAGCAGC	1587			
Sbjct 21360	CCTGTTTGGAAATGAGCCCA	AAATA-----TTCC---GATTGAG---CTCG	21320	Sbjct 20025	TTGGCTGGAGCTGGATGAATGGATGATA---TCTCTCCAGCAGCTGGCGTGCAGCATG	19966			
Query 298	GCTATTGGAAGCTT	CCAGTGAAGAAACCGTATGAATTTTC-TTCAAA-----TACCCTAA	352	Query 1588	AAGGAGTTACGCTGAGCAGCAGCATC-----TTTTGGAGCAAGGAGAACAATTCGGCC	1641			
Sbjct 21319	GCTTTTCAAGAGACTCT	GTAAGGAAACAGAACGAAATTTCTCTCAAAATCTTCCGCTCC	21260	Sbjct 19965	AGGGGATACGCTGAACAGCATCATCAGCAGCTCTTGAAGAACAGCAGCAATCGGCC	19906			
Query 353	TTGTAAATTTGTGTA	ATAA---CTTATGCTAAATTTGGTTCAAAATGTCGAAGAAAGACT	410	Query 1642	TAAACAGCCACCATCGGATCTCCAACTCTATACACAGATCATCCATGGCTTGCACAGCA	1701			
Sbjct 21259	GAGTGATTTGTGTAATAGT	CT---TGCTAAATGTTGTTCTGATATGT---CAAAAAAAGT	21202	Sbjct 19905	TAACTAGCCACCATCGGATCTCCAACTCTATTCAGGCGCTCTACGGCTTACACACCA	19846			
Query 411	TATCCACGCTTCA	TAAGGCTTCACTTAAAGTTGGATT---AGAT-ATTAGGTAA---GGTA	464	Query 1702	GATCAGCACATTGGAATGGGACTGGCAATGGGCTGGGAGGTTCTATCGGTTATCG	1761			
Sbjct 21201	CAACAGCAGCG-CTAG---	CTAC-TGAAATGTAGCTTTAAGGTCATTTGGAAGTTGGGTA	21148	Sbjct 19845	GATCAGCTCAGCTGGAAATGGGATTTGAGGCGGATCATGGCGAGTTCTGTGGTCTATCG	19786			
Query 465	AACACTGTCAACAC	CGGTTTTAGGAGCTTATTATAGAAATCAACATTTAATCAATTT	524	Query 1762	TCCATCTCCAGGACAGGCAAGGAGGACTGTGAGGAAAAATGATGAGGGGATGCAGAG	1821			
Sbjct 21147	AATATTTCAAAAAA	-GGCTATATTAGTTTATATTAATAAGAAAACTTAAGCAATTT	21089	Sbjct 19785	TCCACTCGGAGGACAGGACAGGAGG---TGGACGGCAGGAA-GATGCGCATGGAGAG	19729			
Query 525	GTCCTACGTAAAAA	TTAGGAAATATATAGAAA---GTGCAAGCACCCCTACACATTCGC	583	Query 1822	GAGATTTAGAAAATGAAGCAGATGATGAACGCGACTCGGGAGCTGTGAACAGCTGCTTA	1881			
Sbjct 21088	GTTCAATGCAAAAAAT	GGAAATGTTATAGAAAAGTGCAGAACACAC-TACCCCTTGC	21030	Sbjct 19728	GAGACGGGGATCCAGAGGACAGCATGATCGGACTCTCGCAGCTGAGGCGAGTCTTAA	19669			
Query 584	TTATCCGATGTGT	AAGTACGACCAACACTTACGTGTACCTATGTTTGTGTGTTTACAA	643	Query 1882	GTCAAGCTGTATCAGAGCTGACCTCGGTGAATGATGCGCTATCTCGCCTGGGTTTA	1941			
Sbjct 21029	TCATGAGTATGTG	TAGTGGTACATACACTACGTGTATCTATGTTTGTGTTTACGG	20970	Sbjct 19668	GCCACAGCTGTATCAGAGCTGACCTCGGTGAATGATGCGCTATCTCGCCTGGGTTTA	19609			
Query 644	AGCAACCTTAAAGT	GTGCGACGATGGCAAGCAAAAAAATTTTGTACAACATAATA	703	Query 1942	GTCAAACTCGTATGCTACACTCAACCTTATACCAACTCTCCGCTATATCAACAGTGT	2001			
Sbjct 20969	AGCAACCTTAAAGT	CG-GCAGGATGGCAAGCAAAAAA---TATTTCTGCACTAATAA	20912	Sbjct 19608	GTGACAGACTCTTACGCCACCTTACCCCTTATCAACCACTCTCCGCTATATGACAAATGT	19549			
Query 704	ATAACAATATAAC	GGAAGTACAGATT---GTAAGGCGAGAATTTATTGGAATGTT	762	Query 2002	CCGAGAAGTTCGCTACTCTGGCCACATCTCGGAGGAGACAGTGGAGACACGGATGCA	2061			
Sbjct 20911	CAAA-AAACACGTT	AAAAAGTACATATATAGGAGATG---GAGCTATGTTTAAATGTA	20856	Sbjct 19548	CCGAAAAGTTTCGCTACTCTCGGCCACATCTCTGGAGGAGACAGGCGCAGCGGATGCA	19489			
Query 763	AAGAAACATGTTG	TTTATTTGTTTGTCTATGATTTTATATTAACATAGGATATTAATAA	821	Query 2062	ATGGAGATGGTGCAGGTGGAGGAGTCTGTGAGGTTGTGGAAGTCAACCAACTCACGCG	2121			
Sbjct 20855	TTATATATGCTGT	TAAATTTCAATTTCTGTTTCAATCTGTTTATAATTTGTTAAATTT	20796	Sbjct 19488	ATGTGGAGAGGAGTGGTGGAGGAGTGGTGCAGCGGTGAGAGTTACCAACTCATCCAGCG	19429			
Query 822	-----ACTAGTGAAG	TTATATAAAAAAGACTTATTTCAACGCCTAAA---ATGAATCTA	875	Query 2122	AGGCTACAGGAGTCTTTCAAATATCTAGTGGCAATGCTACATCTCTGTGCTGCCAACA	2181			
Sbjct 20795	TGGACATTTCTTACAG	-ACATAGTTAAGT-TGCTTTGGAAGAGCTGATAGACATTATA	20739	Sbjct 19428	ATGCTCTGGAGCGGTTTAAATCATAGTGGTAAATCCGGCTCTCTCTCTCTCTCTCTCT	19380			
Query 876	TTGACTACCA-TTCT	TTTTTCAAT-TTAAAGC-----ACTTGAGTTAC-----GGT	920	Query 2182	ACGATTGCACTTCTTTTCTGCCCTTCCATGCGCCATAGGAAGTGGCCATTTGGGCTAG	2241			
Sbjct 20738	TAAAAACGACTTCT	TTTAAACGCTTATTTGGTTAACTTAAATTTATTTTATTTAGAT	20679	Sbjct 19379	-----ATTCC-----CTGCCCTTACATGCTCCATGGGAGTGGGCACTTAAGTCTGG	19333			
Query 921	ACC---AACTCAAGC	GTCAATTTGAAGGAATGGATCTCTAAATGATATAATTGACACCC	978	Query 2242	GTGTTTTAAGCGGCTCCAGTCACCATTTTCTCTATACGAAAAGCTATCTTCAATGATT	2301			
Sbjct 20678	AACGTAATCCAGAG	ACATCTTG-AGGGGATGGAGTCTAATAGTGAATTAATTGACACCC	20620	Sbjct 19332	GGGTGTTGAGCGGTTGACAACTCTCATATTTCTCATATGAGAGATCTGCTCAATGATT	19273			
Query 979	AAACATTTAGCCAG	GAGTTAGTTGAAGATGATCGGAGTTCATCACTGGGTCATCATTT	1038	Query 2302	CCCCCGGCTAAATAACTATTTGGTCTGCTGGGATCTACATCTCTCGTTTCCGGAACCG	2361			
Sbjct 20619	AGACTTTTAGCCAGG	ATCTGGTTGAAGATGCAACAGATTTTATCTCTGTTGGGCGCAACT	20560	Sbjct 19272	CACCCCAACCAATAGCTACTTGGTCTGCTGGGATCTGATGCTGCTGCTGCTGCTGCTG	19213			
Query 1039	CGGAGCGCCCTTC	GCACTC---CAG-----TCAGCAG---CCTAATCTGCGACAGG	1083	Query 2362	TTATAAATTCATCGCACTTGCAGCTAAGTCAACAGGCAATAAAGGAGTCTGGAGCA-	2420			
Sbjct 20559	CGGAGCGCTCAGT	CTCATCCGCACTGCAACATCAGAGGGGAGAGAACCTGATTCGCGCGAGG	20500	Sbjct 19212	TTCTTAATCATCTCAACTGACGCTAAACACACAGCGGCAAAAAAAGAACTAGCAATAG	19153			
Query 1084	ATCTGACAAATGT	CTATGCAAGACATAATTTCTGTCCG-----GTG---	1124	Query 2421	---CACGAACACACATAGGCTGCTGATGTCAATGTTGGGAGGTTTTCTTATATCTGGCC	2478			
Sbjct 20499	ACCTGGCGATTTCT	CCCTGCAACGATGACTCTCTGCCAAGAACGGACGGAGTACTAGTGAG	20440	Sbjct 19152	CTCACGAACACCCCATGGTATGCTGCGATGCTCAATGGGCGGAAGTTCTCTACACGGCC	19093			
Query 1125	-----AAACACGG	AGCTGCTGCTATCTGGGTCGGGAT-----CGGATCGGGATCCG	1173	Query 2479	ACATCTCCGAGGTGATAGTGTAGATAATGACGTCAACGGTGAAGAGTTCTTCTTCTCTG	2538			
Sbjct 20439	GGAAGAAACACCG	GCTGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTG	20380	Sbjct 19092	ACATATCTGGAGGAGATGTGAGATACCGATGCTCAATGGGAGGAGTTTCTTACTCTG	19033			
Query 1174	ATTCTGTTGTCAG	TGTTATAGACGCACTGGGCAAGGTAACAGGCACTGACATATCAA	1233	Query 2539	ACCATATCTCCGGTGGGATAGCGGAGATGAAGTCCACAGAGAGAAAAATTCATATACT	2598			
Sbjct 20379	ATTCTGTTGTCAGT	GTATTGCACTCAATGGGCAAGCAATGACAGCAACATTTCAA	20320	Sbjct 19032	ATCACATCTCTGGAGGAGACAGCGGAGCAGGATGCTCAATGAGAGAAATTTCTTACTCT	18973			
Query 1234	TTTGTGCTCAGCA	GCTGCAAGAAACATGCACTTGGCTTGGGCTCTGGAACAG	1293	Query 2599	CAGATCAGATCTTCCGAAGGAGAAACGGACCGGAGCTCAATAGCGGAAC	2647			
Sbjct 20319	TTTGTGCTCAGCA	GATGATCCAGAAA---TGCCATTGCCCTCGGGTGTCTCCAGCAA	20263	Sbjct 18972	CTGATCAGATCTCTGGAGCAGAAAGCGTTCGGGTGCTCAACGGAGGAAC	18924			

2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (e.g., TopHat, regtools), **paste a Genome Browser screenshot of the region surrounding the putative TSS (± 300 bp) with the following evidence tracks:**

1. RNA-Seq Alignment Summary or RNA-Seq Coverage
2. RNA-Seq TopHat or Splice Junctions

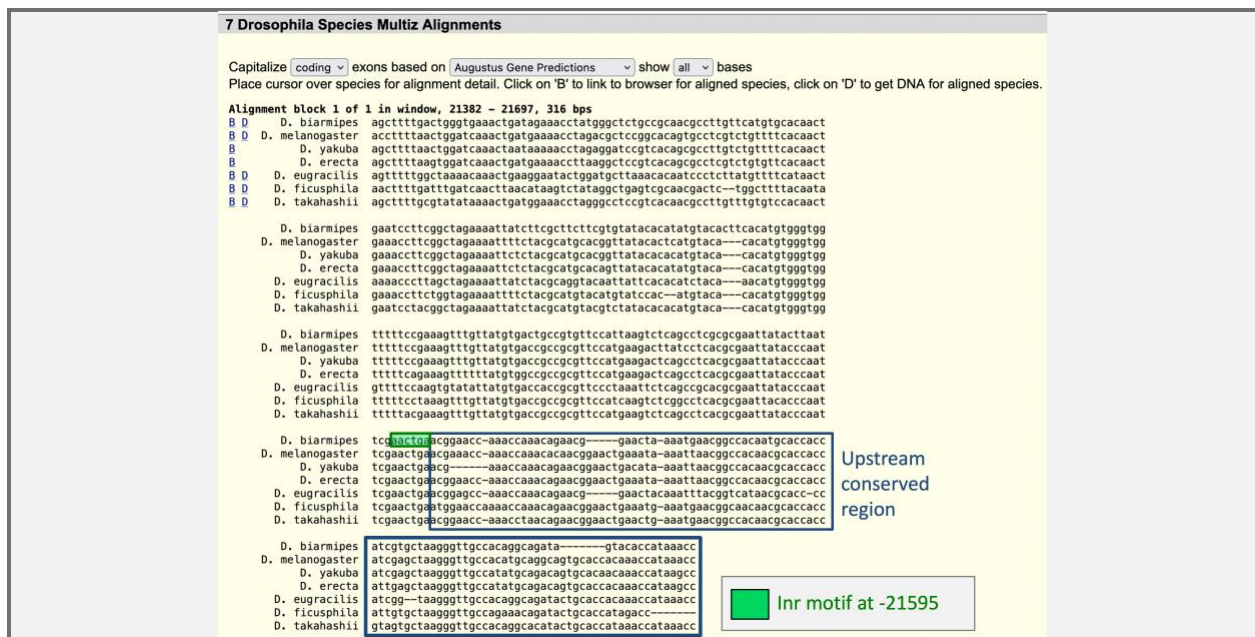


If the RNA-Seq evidence tracks indicate the TSS position, list it here: NA (see explanation below)

If the RNA-Seq evidence tracks indicate a TSS search region, list it here: contig35:21,599–21,750

3. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, paste a screenshot of the multiple sequence alignment (e.g., from Clustal Omega, ROAST) into the box below:



4. Define the TSS search region(s)

Note:

If the *blastn* alignment to the initial transcribed exon satisfies the criteria listed on page 8 of Module TSS4 (i.e., a long match with low E-value, requires extrapolation of less than 150 bp to the estimated TSS position, alignment is in concordance with other evidence tracks), then you can define the TSS search region as +/- 300 bp from the initial 5' nucleotide. For example, if the estimated TSS position is located at position 1500, then the narrow TSS search region would be placed at 1200-1800.

If you cannot estimate the TSS position based on the *blastn* alignment to the initial transcribed exon, then you can define the TSS search region(s) based on the experimental data (e.g., RNA-Seq, RNA PolII ChIP-Seq) and the conservation track for the target species. If part of the TSS search region is only weakly supported by the available evidence, then please specify both a **wide** and a **narrow** search region. For example, if the region at 1500-2000 shows high RNA-Seq read coverage but there is very low RNA-Seq coverage from 1000-1499, then you will report “**1000-2000**” as the wide search region and “**1500-2000**” as the narrow search region.

Enter “Insufficient evidence” if the narrow search region cannot be determined based on the available evidence.

Coordinates of the narrow TSS search region:

21,599–21,750

Coordinates of the wide TSS search region:

(Enter “NA” if the narrow TSS search region is defined based on the *blastn* alignment to the initial transcribed exon. Enter “Insufficient evidence” if a wide search region cannot be defined based on the available evidence)

Insufficient evidence

Describe the evidence used to define the TSS search region(s) (e.g., RNA-Seq and Conservation tracks in this species, RAMPAGE data from *D. melanogaster*):

The TSS search region is defined based on RNA PolII enrichment, RNA-Seq read coverage, sequence conservation among seven *Drosophila* species, and the Inr motif at 21,630 (see the explanations in section 6 below).

5. Search for core promoter motifs

The consensus sequences for the *Drosophila* core promoter motifs are available at https://gander.wustl.edu/~wilson/core_promoter_motifs.html

Use the "Short Match" functionality in the GEP UCSC Genome Browser to search for each of the core promoter motifs listed below **in the region surrounding the TSS (± 300 bp) in your project and in the *D. melanogaster* ortholog.**

For TSS annotations where you can only define a TSS search region (and not a single coordinate), you should report all motif instances within the narrow TSS search region. If you did not report a narrow TSS search region due to insufficient evidence, report the motif instances in the wide TSS search region.

Coordinates of the motif search region

Your project (e.g., contig10:1500-2000): contig35:21,299-21,899

Orthologous region in *D. melanogaster*: chr4:607,349-607,949

Record the **orientation and the start coordinate** (e.g., +10000) of each motif match below. (Enter "NA" if there are no motif instances within the search region.)

Note: Highlight (in yellow) the motif instances that support the TSS annotation above.

Core promoter motif	Your project	<i>D. melanogaster</i>
BRE ^a	NA	NA
TATA Box	NA	NA
BRE ^d	-21409, -21494, -21606, -21611, -21640, -21650, -21686, -21695, -21818	-607557, -607635, -607656, -607661, -607663, -607666, -607695, -607705, -607745, -607748, -607757
Inr	-21399, -21448, -21595 , -21630, -21883	-607452, -607501, -607645 , -607675
MTE	NA	NA
DPE	-21361, -21446, -21809	-607499, -607831, -607905
Ohler_motif1	NA	NA
DRE	NA	NA
Ohler_motif5	NA	NA
Ohler_motif6	NA	NA
Ohler_motif7	NA	NA
Ohler_motif8	NA	NA

6. Summarize all of the evidence that supports the TSS annotation postulated above.

Coordinate(s) of the TSS position(s):

Based on *blastn* alignment: 21,599Based on core promoter motifs (e.g., Inr): 21,598Based on other evidence (please specify): NA

Note: If the *blastn* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter “Insufficient evidence” if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence? Yes; at 21,599

If so, indicate whether the evidence listed below support the TSS position.

If not, indicate whether the evidence listed below support the TSS search region(s).

Evidence type	Support	Refute	Neither
<i>blastn</i> alignment of the initial exon from <i>D. melanogaster</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RNA PolII ChIP-Seq	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
RNA-Seq coverage and splice junctions	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Core promoter motifs	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sequence conservation with other <i>Drosophila</i> species (e.g., “Conservation” track on the Genome Browser)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
RefSeq Genes, N-SCAN PASA-EST, and Augustus TSS predictions	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Other (please specify) TSS identified by the RAMPAGE datasets in <i>D. melanogaster</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

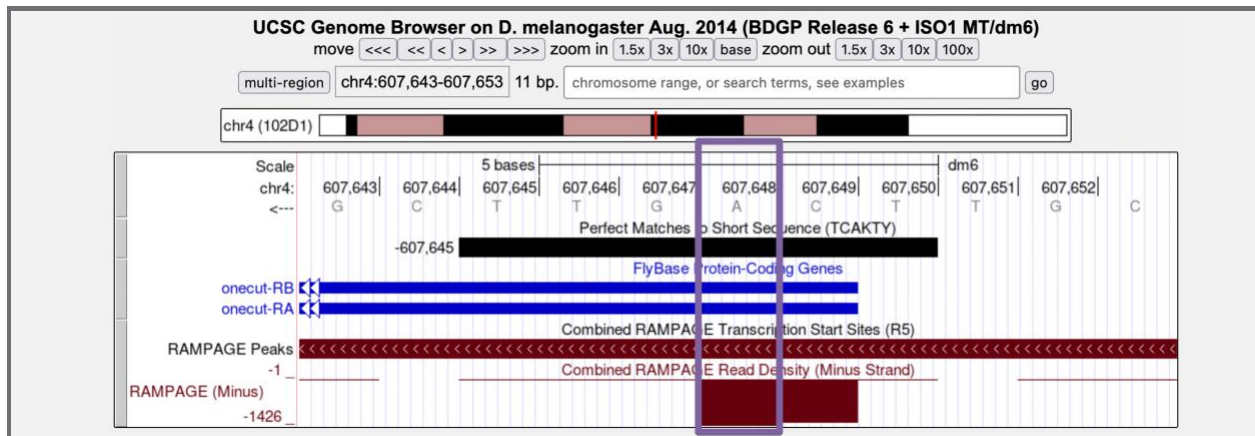
Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against (i.e., Refute) the annotated TSS. In contrast, the lack of RNA-Seq read coverage is a negative result and it neither supports nor refutes the TSS annotation (i.e., Neither).

Provide an explanation if the TSS annotation is inconsistent with at least one of the evidence types specified above:

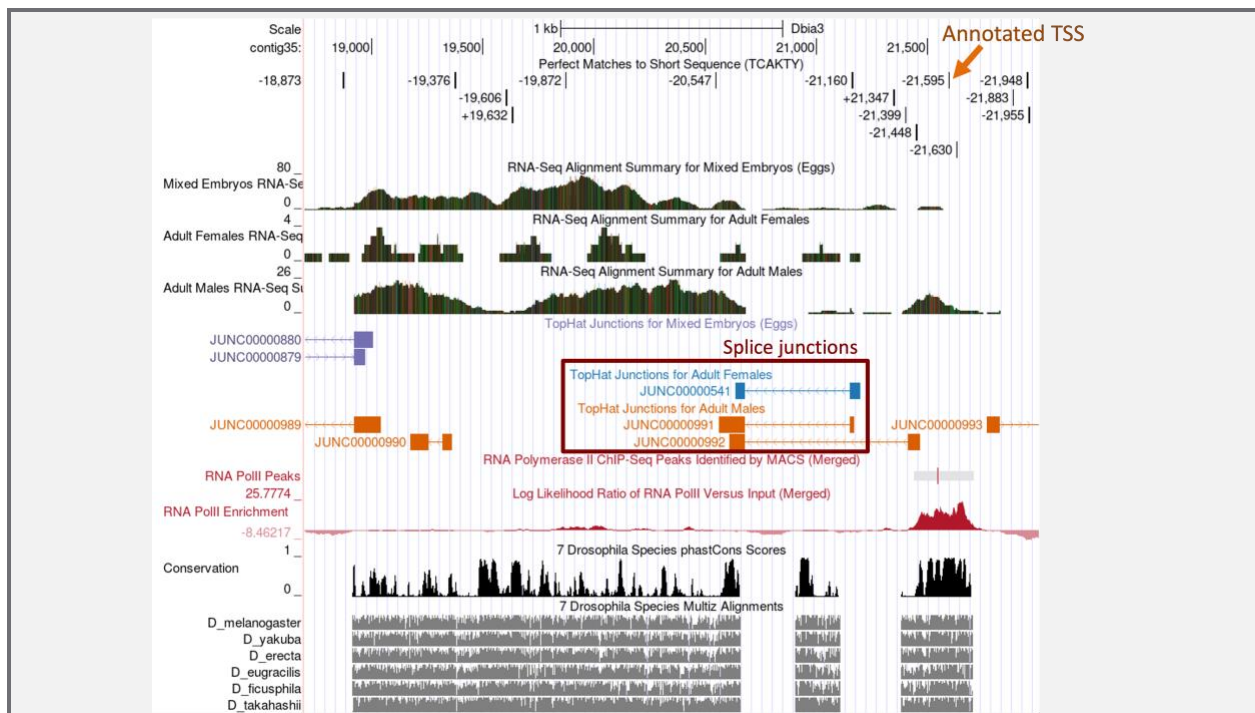
The annotation of the TSS position for the *D. biarmipes onecut* gene is based on minimizing the change in the size of the initial transcribed exon compared to the *D. melanogaster* ortholog (i.e., parsimony). The *blastn* alignment of the initial transcribed exon of the A isoform of *onecut* (*onecut:3*) from *D. melanogaster* against the genomic sequence of *D. biarmipes* contig35 shows a full-length alignment and the exon is highly conserved (with an E-value of 0.0 and a sequence identity of 71%; see screenshot above). The *blastn* alignment placed the TSS of *onecut* at 21,599. There is also an Inr

motif at 21,595–21,600. Because the Inr motif is found at -2 of the TSS, the Inr motif would place the TSS of *onecut* at 21,598.

Examination of the Combined RAMPAGE TSS datasets for *D. melanogaster* suggest there are two strong TSS in *onecut*. The TSS at 607,649 is consistent with the *D. melanogaster* FlyBase annotations for the A and B isoforms of *onecut* while the TSS at 607,648 is consistent with the Inr motif. Both TSS have high RAMPAGE read density but the TSS at 607,648 is stronger than the TSS at 607,649. Hence the TSS of *onecut* for *D. biarmipes* is placed at 21,599 based on parsimony with *D. melanogaster*.

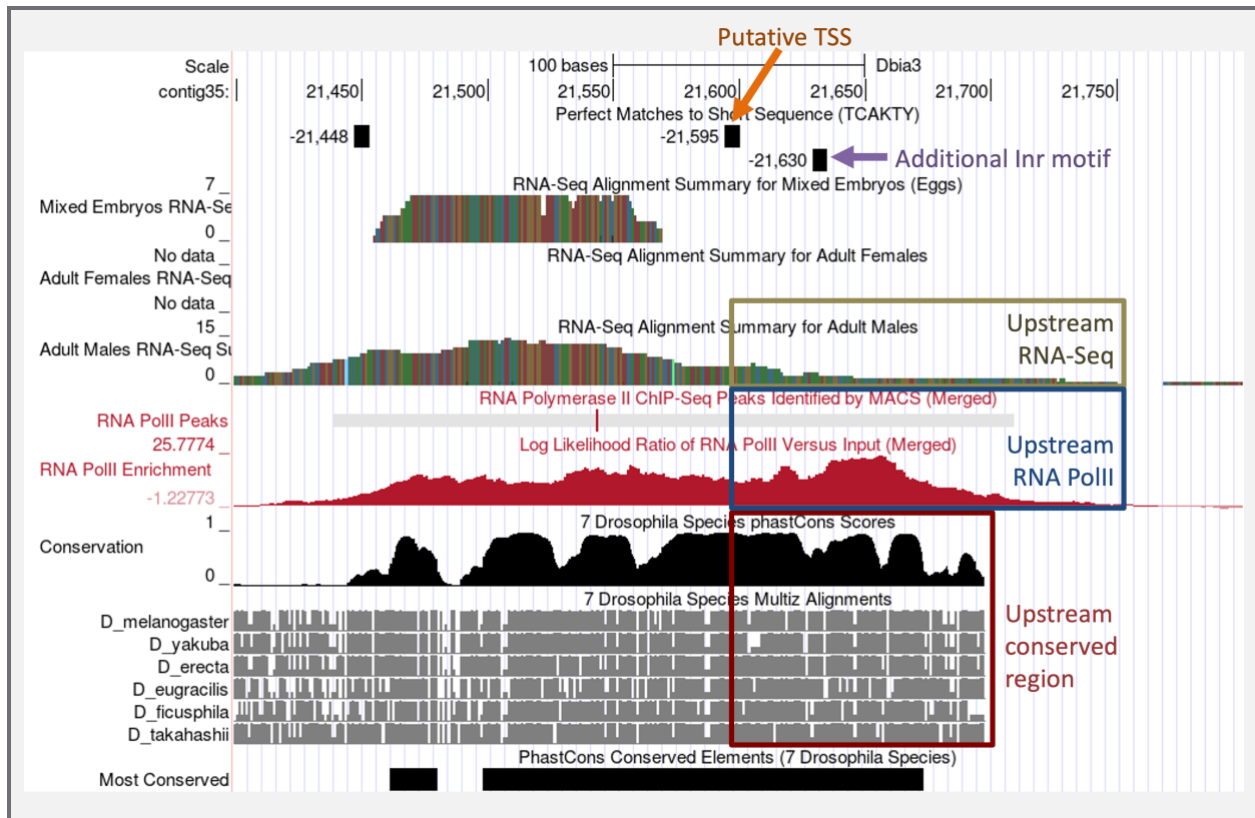


Examination of the region surrounding the *blastn* alignment to exon onecut:3 (i.e., 21,599–18,924) shows that the available RNA-Seq data generally supports the proposed TSS position at 21,599. For example, the TopHat junctions in adult females and adult males (JUNC00000541 and JUNC00000991, respectively) are consistent with the splice junction between exons onecut:1 and onecut:2 in the B isoform of *onecut*.

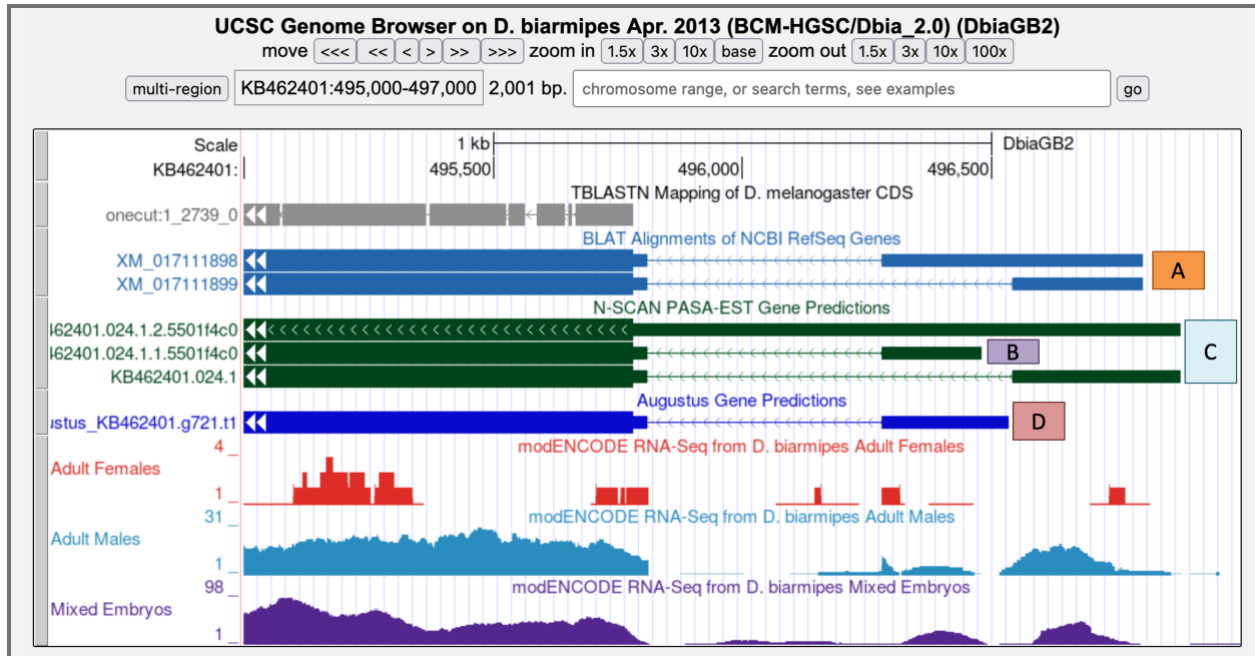


While the RNA-Seq evidence is generally in congruence with the proposed TSS position, the RNA-Seq alignment summary from the adult males sample shows additional RNA-Seq read coverage upstream (i.e., further to the right) of the annotated TSS.

Similarly, the "RNA PolII Peaks" and the "RNA PolII Enrichment" tracks indicate the presence of RNA PolII upstream of the annotated TSS at 21,599. This region also has an Inr motif at 21,630. According to the "Conservation" and the "Most Conserved" tracks, part of this upstream region (21,600–21,697) is highly conserved with the orthologous regions from six *Drosophila* species (*D. melanogaster*, *D. yakuba*, *D. erecta*, *D. ficusphila*, *D. eugracilis*, and *D. takahashii*). Collectively, the available evidence suggests that the TSS of *onecut* might be located further upstream of the proposed TSS for *onecut* in contig35.



Examination of the RefSeq genes, N-SCAN, and Augustus gene predictions in the *D. biarmipes* April 2013 (BCM-HGSC/Dbia_2.0) assembly shows that each tool assigned different position(s) as the TSSs for *onecut* (labeled A–D in the figure below). Due to these inconsistencies and the high error rate associated with TSS predictions, these TSS predictions were not used to define the TSS position or the TSS search region for the *onecut-RA* ortholog in *D. biarmipes*.



Collectively, the narrow TSS search region for *onecut-RA* is defined at contig35:21,599–21,750 to account for the RNA-Seq read coverage, RNA PolII ChIP-Seq data, conservation among seven *Drosophila* species, and the location of the additional Inr motif.