

The Complexity of eIF4E1, Its Impact on Ortholog Identification, and Improvements for Student Success with Gene Annotation

Katherine Free, Dr. Jack Vincent
Biomedical Science, University of Washington Tacoma

BACKGROUND

The Genomics Education Partnership (GEP) Pathways Project uses a network-based analysis to understand the function of metabolic and signaling pathways and their evolution across the genus *Drosophila*. Currently, the Pathways Project is using a pipeline method that allows undergraduate students to complete annotations on genes located in the insulin signaling pathway of the *Drosophila* species.

SIGNIFICANCE

Gene Annotation

- Recognize how regulatory regions of particular genes evolve within a network of a species

The Eukaryotic Translation Initiation Factor 4E1, eIF4E1

- Crucial component of the *Drosophila* genome

Homologs	Sequence Location
eIF4E1	3L
eIF4EHP	3R
eIF4E3	3L
eIF4E4	3L
eIF4E5	3L
eIF4E6	3R
eIF4E7	X

Figure 1. Table showing the seven homologs of the gene eIF4E1. These homologs make identifying the ortholog difficult while annotating.

OBJECTIVE

- Annotate eIF4E1 in *D. obscura*, *D. navojoa*, *D. takahashii* following the Pathways Project Walkthrough step-by-step
- Examine the *tblastn* results
- Compare genomic neighborhoods with the reference species
- Determine if the ortholog is identified

HYPOTHESIS

The eIF4E1 gene will not be able to be annotated in *D. obscura*, *D. navojoa*, or *D. takahashii* using the current Pathways Project Walkthrough criteria. Identifying the ortholog rather than one of the multiple homologs will be challenging when looking for the *tblastn* result that has both the highest percent identity and the lowest E value.

METHODS AND MATERIALS

- Examine *tblastn* search results from the protein sequence of the eIF4E1 gene in *D. melanogaster*
- Look for the *tblastn* result that best fits the walkthrough criteria of highest percent identity and lowest E value
- Record approximate coordinates of the best fitting scaffold
- Compare the genomic neighborhood of the target species to *D. melanogaster*
 - If neighborhoods are similar, the ortholog has been identified
 - If neighborhoods are not similar, a homolog has been identified

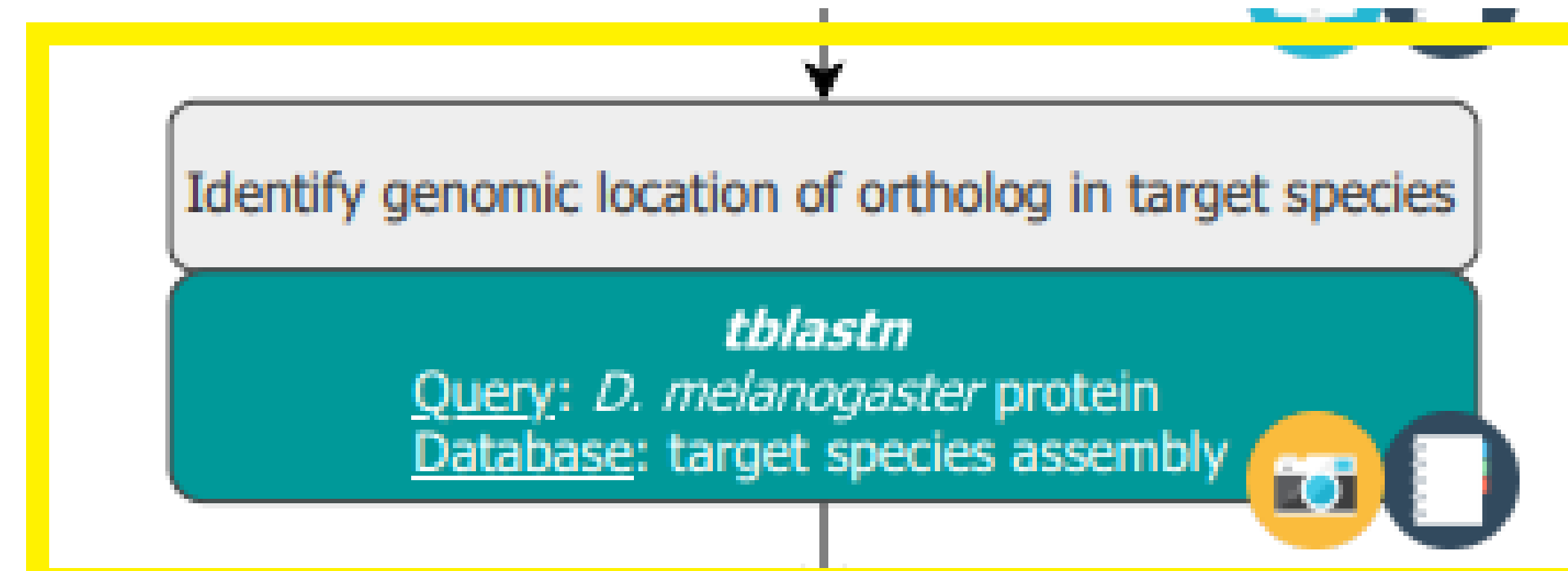


Figure 2. Part 2 of the Pathways Project Walkthrough. This is the most relevant step to my research as this is where ortholog identification is first required.

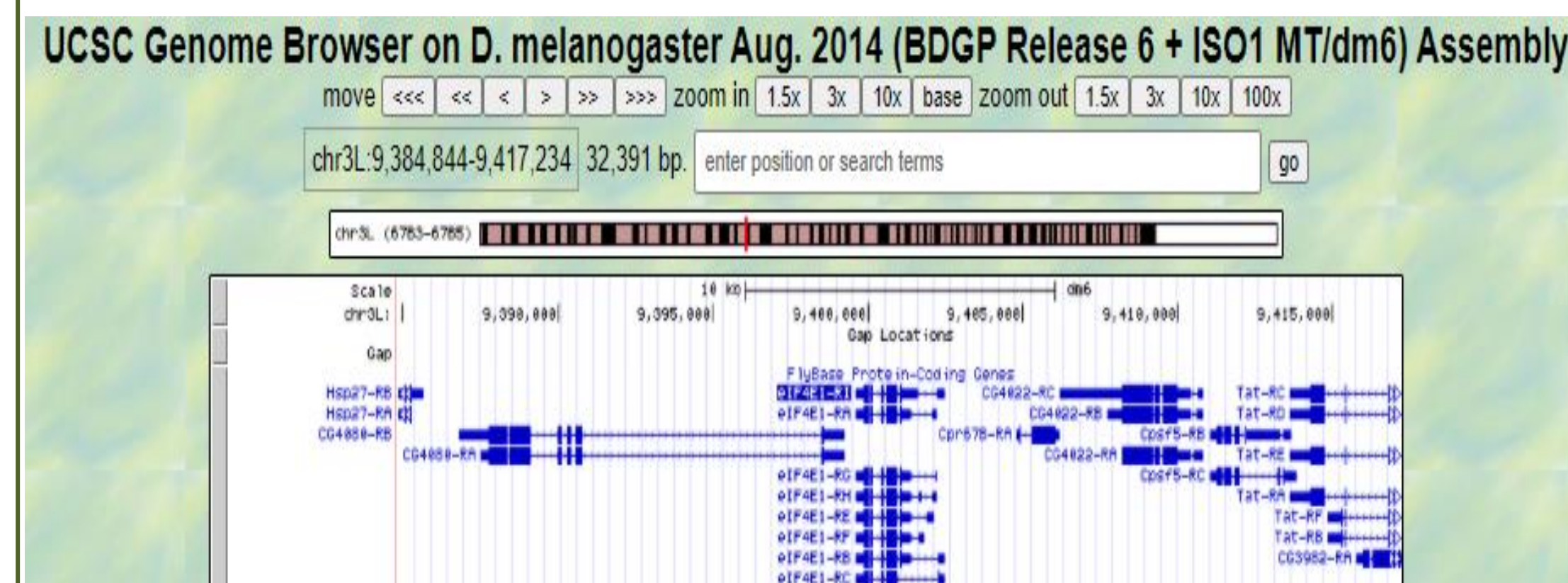


Figure 3. The genomic neighborhood of eIF4E1 in *D. melanogaster*.



Figure 4. Simplified version of the genomic neighborhood of *D. melanogaster* showing eIF4E1 and its surrounding genes. Any ortholog of eIF4E1 will have these genes present in its genomic neighborhood. This version was used for comparative purposes.

RESULTS

Description	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
Drosophila takahashii Contig6705, whole genome shotgun sequence	259	259	69%	2e-78	66.11%	127920	AFFI02006708.1
Drosophila takahashii unplaced genomic scaffold scf7180000415664, whole genome shotgun sequence	187	392	94%	1e-95	60.23%	438030	KB461488.1
Drosophila takahashii Contig5696, whole genome shotgun sequence	75.1	75.1	45%	2e-14	32.50%	134607	AFFI02005701.1
Drosophila takahashii Contig9043, whole genome shotgun sequence	57.4	163	69%	8e-28	64.86%	110917	AFFI02009046.1
Drosophila takahashii Contig7317, whole genome shotgun sequence	103	250	69%	3e-24	72.31%	268029	AFFI02007320.1
Drosophila takahashii Contig8779, whole genome shotgun sequence	127	288	64%	2e-32	89.06%	583212	AFFI02008782.1
Drosophila takahashii Contig8283, whole genome shotgun sequence	181	181	67%	4e-51	52.54%	440458	AFFI02008286.1
Drosophila takahashii Contig8184, whole genome shotgun sequence	187	392	94%	2e-95	60.23%	306224	AFFI02008187.1
Drosophila takahashii unplaced genomic scaffold scf7180000415700, whole genome shotgun sequence	181	181	67%	4e-51	52.54%	573062	KB461517.1
Drosophila takahashii unplaced genomic scaffold scf7180000415809, whole genome shotgun sequence	127	288	64%	2e-32	89.06%	1234928	KB461809.1
Drosophila takahashii unplaced genomic scaffold scf7180000415352, whole genome shotgun sequence	103	250	69%	3e-24	72.31%	1175522	KB461236.1
Drosophila takahashii unplaced genomic scaffold scf7180000415862, whole genome shotgun sequence	57.4	163	69%	6e-28	64.86%	447564	KB461860.1
Drosophila takahashii unplaced genomic scaffold scf7180000414373, whole genome shotgun sequence	75.1	75.1	45%	2e-14	32.50%	229605	KB460892.1

Figure 5. Results of the *tblastn* search for *D. takahashii*. Results #2 and #6 are highlighted. Result #6 most closely matches the walkthrough criteria, while Result #2 is the actual scaffold for the ortholog of eIF4E1.

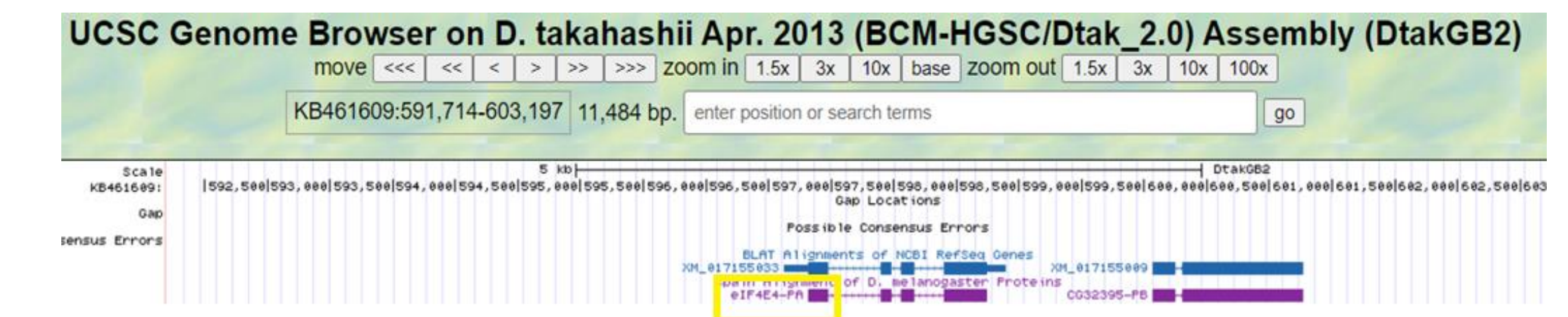


Figure 6. The genomic neighborhood of the *tblastn* result that has the highest percent identity and lowest E value, Result #6 (Fig.5). This is the best match to the Pathways Project Walkthrough criteria, but shows the homolog eIF4E4 rather than the ortholog.

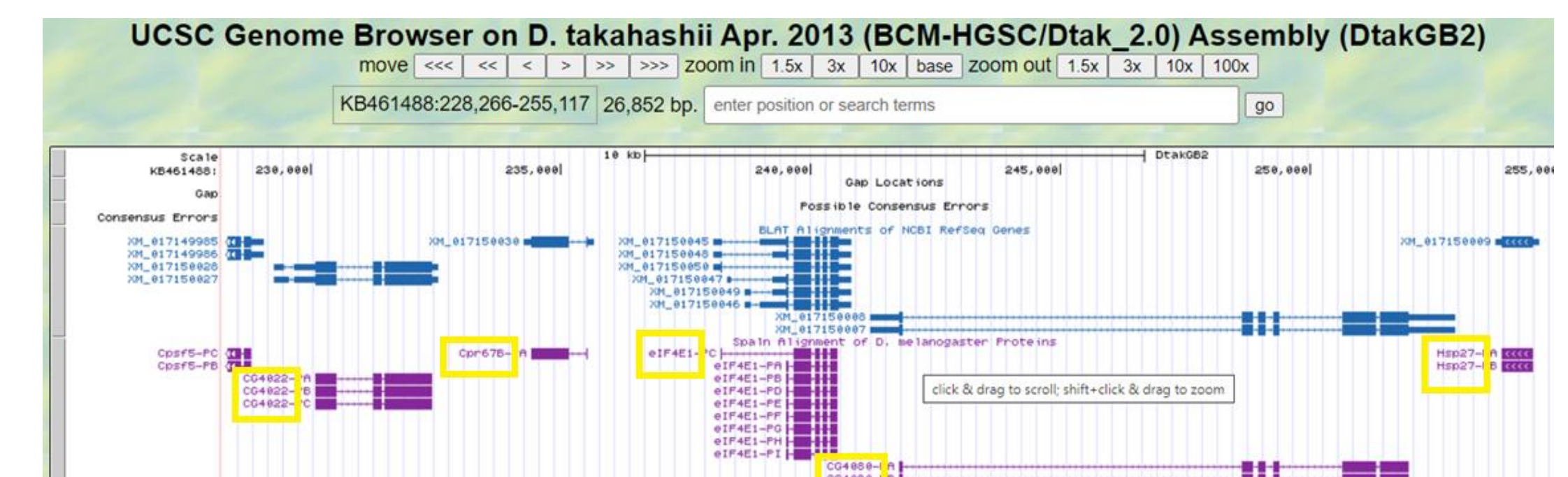


Figure 7. The genomic neighborhood of the *tblastn* Result #2. This result does not meet the criteria of the highest percent identity and lowest E value. However, this scaffold is the ortholog eIF4E1, proven by the appearance of the same genes in *D. melanogaster*.

CONCLUSION

Our hypothesis is supported. The gene eIF4E1 cannot be annotated using the Pathways Project walkthrough as written in *D. obscura*, *D. navojoa*, or *D. takahashii* due to the presence of multiple homologs. We have proposed to the GEP that looking at the genomic neighborhood before identifying potential coordinates of a gene location would be the first step in adjusting the annotation protocol to accommodate these issues with the eIF4E1 gene in *Drosophila* species.

REFERENCES

- Genomics Education Partnership (GEP) Mission. 2017. Tuscaloosa (AL): The University of Alabama Genomics Education Partnership <https://thegep.org/aboutsite/>
- Sandlin K, Leung W, Reed L. 2021. Pathways Project: Annotation Walkthrough. Genomics Education Partnership.