The hunt for the non-canonical start site in the *FMR-1* gene

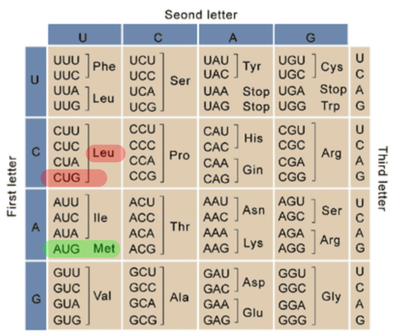
Dionne Weaver, Gianni Walker, and Dr. Jack Vincent

UNIVERSITY of WASHINGTON TACOMA

Introduction:

While working on a previous annotation project, evidence for a non-canonical (or non-AUG) start site was found in two isoforms for the *FMR-1* gene in the fruit fly species, *Drosophila erecta.* Upon further research, we found a paper that discussed the conservation of the 5' UTR within the *melanogaster* subgroup. We then wondered if this conservation also meant the non-canonical start site was conserved and if so, how far into the *Drosophila* genus it went.

This project looked for evidence a non-canonical start site within the same isoforms of four additional species. The species looked at were *D. ficusphila, D. obscura, D. arizonae, and D. buskii* and were chosen based on their evolutionary split from *D. melanogaster* whose genome has been annotated.



Green: typical AUG start codon

Red: non-canonical start codon found within the G and H isoform of *D. melanogaster*

For more information please contact:

DionneW@UW.EDU, GWalker9@UW.EDU, or JaV2@UW.EDU

Objectives:

Our objectives for this project were to 1. annotate the *FMR-1* gene in the four chosen species and 2. analyze the G and H isoforms for evidence of the non-canonical start site.

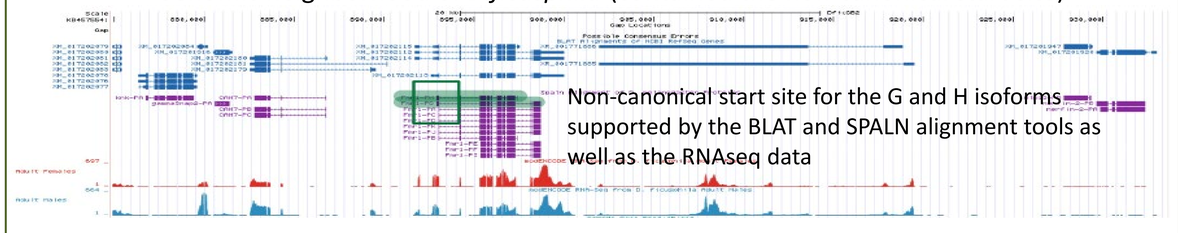
Materials & Methods:

The GEP UCSC genome browser was first used to examine the genomic neighborhood of *D. melanogaster.* The predicted protein sequence for the *FMR-1* gene was used as the query for the tblastn search to find the ortholog in the four chosen species. Once the ortholog was found, the genomic neighborhood was examined. Using the accession numbers from the refseq BLAST alignment we obtained the predicted protein sequences and compared them to that of *D. melanogaster*. After visual confirmation of the ortholog, the gene record finder was used to find the approximate locations of the coding exons (CDS's). The location of the CDS's was then refined using a BLAST search comparing the sequence within *D. melanogaster* to the target species. Using the genome browser, we then visually examined each exon for evidence of the non-canonical start sites, focusing on the G and H isoforms. We looked for any strong indications of homology between the BLAT, SPALN, and refseq data. The potential non-canonical start site was confirmed by both the CUG codon as well as support from RNAseq data.

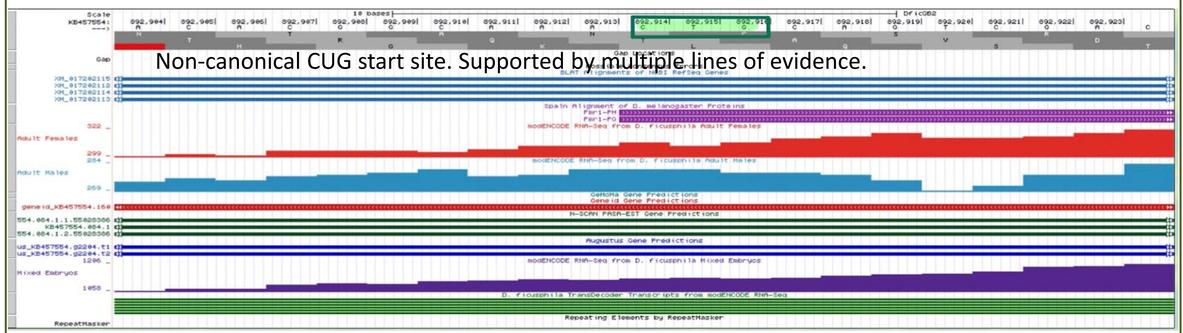
Results:

Of the four species examined, only one species, *D. ficusphila* showed evidence of the non-canonical start site.

Genomic neighborhood of *D.ficusphila* (contained the non-canonical start site)



Non-canonical start site for the G and H isoforms supported by the BLAT and SPALN alignment tools as well as the RNAseq data

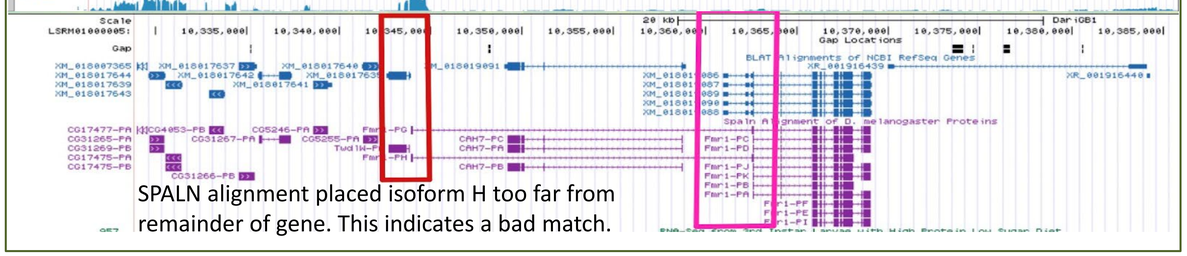


Non-canonical CUG start site. Supported by multiple lines of evidence.

Genomic neighborhoods of *D. obscura, D. buskii.* and *D. arizonae* (did not contain the non-canonical start site)

RNAseq data does not support transcription in this area. This indicates a bad match

We would not expect to see genes in between the exons of the *FMR-1* gene. This indicates a poor match for the G and H isoforms



SPALN alignment placed isoform H too far from remainder of gene. This indicates a bad match.

References:

Beerman RW, Jongens TA. 2011. A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms. Neuroscience. 181:48–66.

Acknowledgements:

Genome pathways project for giving us a research topic

University of Washington Tacoma Sciences department for the opportunity to do a research project