

AgAnimalGenomes: Browsers for Viewing and Manually Annotating Farm Animal Genomes

Mammalian Genome

Deborah A. Triant, Amy T. Walsh, Gabrielle A. Hartley, Bruna Petry, Morgan R. Stegemiller, Benjamin M. Nelson, Makenna M. McKendrick, Emily P. Fuller, Noelle E. Cockett, James E. Koltes, Stephanie D. McKay, Jonathan A. Green, Brenda M. Murdoch, Darren E. Hagen, Christine G. Elsik

Corresponding author:

Christine Elsik

Division of Animal Sciences, University of Missouri, Columbia, MO 65201, USA

elsikc@missouri.edu

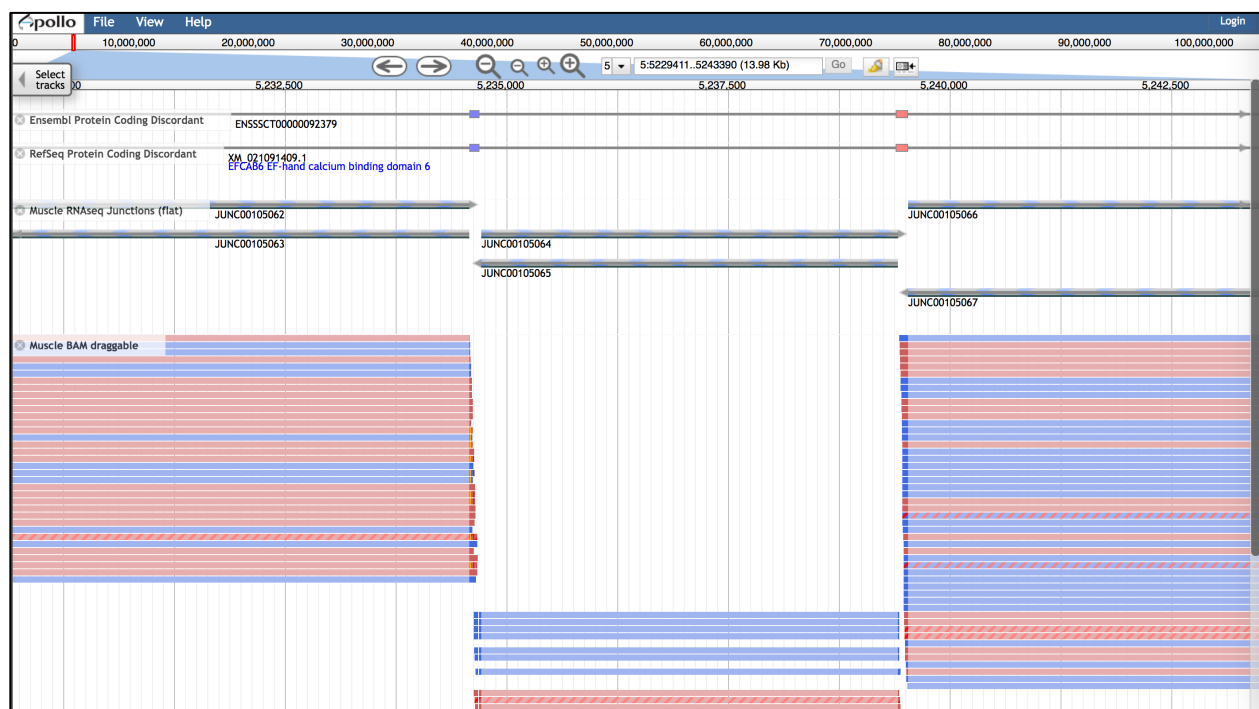


Fig. S1. RNAseq tracks shown as flat junctions and read alignments (BAM) in the draggable format. The zoom level must be sufficient to avoid the “Too much data to show” error. Blue and red colors in the BAM track represent the strand of the read alignment. Spans between aligned read segments are shown in lighter red and blue. The darker blue and red aligned portions of the reads are barely visible. The reads with hash marks are paired-end reads with a missing paired read.

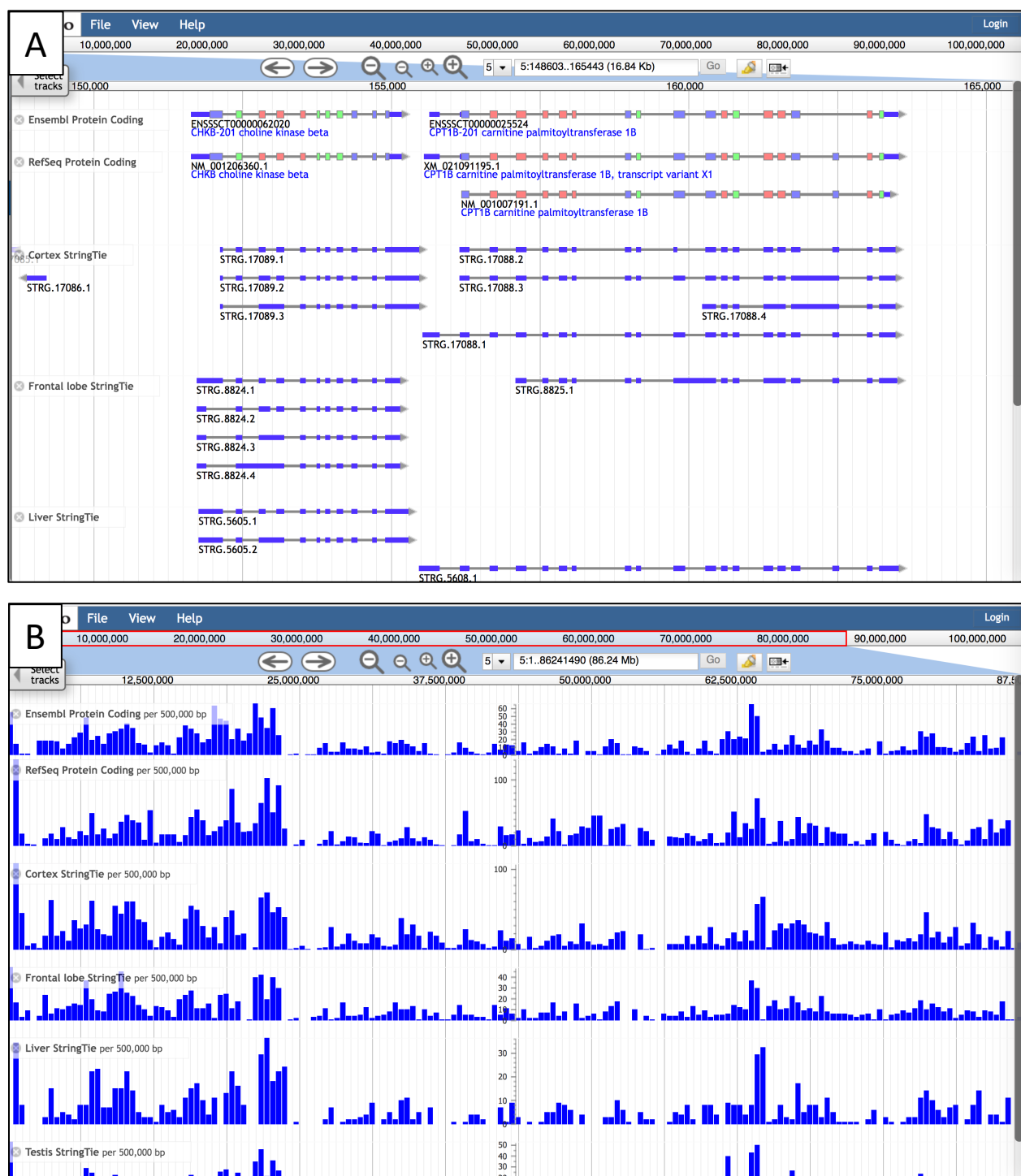


Fig. S2. A) StringTie tracks, zoomed-in. Notice all the exons in StringTie tracks are shown as dark blue (noncoding RNA) because the coding potential is not known. B) StringTie tracks zoomed out.

A

SequenceServer 2.0.0 [Help & Support](#)

```
>XM_005677499.3
AGAGGGAGTTCAGTGAACCTCCACCAGGAGCTCCGCCGGGAGGGCAGGAGGG
AGTGGGACAGAATCTGGGAGTGCAGCTGGGAGGAGTCTGGCTGGGCTGAGCGGGGAG
CTGCTTGGCAGTGCAGAGCCAGGCCCCAGGAGGAGGAGTGGACCGAGA
AGGCAGGCCCTGTGAGGAGTGCAGTGGGAGGAGTCTACAGGAGCTAGAGAAGCGGTG
GTGCTCTGTGGGAACTTCTACAAATATGTCTCAAGCAGCAGCTGTGCAAGAACAAG
ATCAGCAAGAGCAGCTTCCGGAAGATGCTTCCGAAAGAGCTCAACCATATGCTGACGGAC
ACGGGGAACCGGAAGCTGCTGACAACTCATCCAGGAGCTGAGCGCCACCAAGCAGGGA
CGCATCAGCTTTGACGAGTACTGGACCTTGATAGGCGGATCATAGTCCATATGCCAAC
CTTATTCGCGCAGCAGGAGCAGACAGCAGCTAGAGCCCTCCACCCCTCTTCCCC
CCATGCCGCTCTAGGCTCCCTCTCTGCCCCAGGAGGCCCTTCTCCCTCTCTCC
CTCTAGACCTCTCTGGACCATAGCTGAACAGGATGAGGGAACGTCAGGCCCCCTTCG
```

Nucleotide databases [Select all]

- ☐ Bos taurus ARS-UCD1.2 genome
- ☐ Bubalus bubalis NDDB SH 1 genome
- ☒ Capra hircus ARS1 genome
- ☐ Equus caballus EquCab3.0 genome
- ☐ Gallus gallus GRCg6a genome
- ☐ Ovis aries ARS-UI Ramb v2.0 genome
- ☐ Sus scrofa Sscrofa11.1 genome

Advanced parameters: ? ☒ Open results in new tab **BLASTN**

B

BLASTN: 1 query, 1 database
[Edit search](#) | [New search](#)

Download FASTA, XML, TSV
[FASTA of all hits](#)
[FASTA of selected hit\(s\)](#)
[Alignment of all hits](#)
[Alignment of selected hit\(s\)](#)
[Standard tabular report](#)
[Full tabular report](#)
[Full XML report](#)

SequenceServer 2.0.0 using **BLASTN 2.12.0+**, query submitted on 2023-04-23 22:04:03 UTC
Databases: Capra hircus ARS1 genome (29907 sequences, 2922813246 characters)
Parameters: evaluate 1e-20, sc-match 1, sc-mismatch -2, gap-open 0, gap-extend 0, filter L;m;
Please cite: <https://doi.org/10.1093/molbev/msz185>

Query= goat_cdna length: 1,098
[Graphical overview of hits](#) [SVG](#) [PNG](#)

☐ Length distribution of matching sequences
☐ Sequences producing significant alignments

#	Similar sequences	Query coverage (%)	Total score	E value	Identity (%)
1.	gnl ARS1 3 CM004564.1 NC_030810.1 Chr3	100	1104	0	100%

gnl|ARS1|3 CM004564.1 NC_030810.1 Chr3 length: 120,038,259
[Select](#) | [Sequence](#) | [FASTA](#) | [Alignment](#) | [View in JBrowse](#) [SVG](#) [PNG](#)

☐ Graphical overview of aligning region(s)

a. Score: 1371.33 (742), E value: 0, Identity: 742/742 (100%), Gaps: 0/742 (0%), Strand: + / -
Query 357 GGACACGGGGAACCGAAAGCTGCTGACAAGCTCATCCAGGACCTGGACGCCAACACAGA 416
|||||
Subject 103378418 GGACACGGGGAACCGAAAGCTGCTGACAAGCTCATCCAGGACCTGGACGCCAACACAGA 103378359
Query 417 CGGACGCATCAGCTTTGACGAGTACTGGACCTTGATAGGCGGCATCACTAGTCCTATCGC 476
|||||
Subject 103378358 CGGACGCATCAGCTTTGACGAGTACTGGACCTTGATAGGCGGCATCACTAGTCCTATCGC 103378299

C

File View Help Login

10,000,000 20,000,000 30,000,000 40,000,000 50,000,000 60,000,000 70,000,000 80,000,000 90,000,000 100,000,000 110,000,000 120,000,000

Select tracks

blasttmp1206780_103378418

blasttmp1206780_103378892

RefSeq Protein Coding

XM_018046138.1
S100A16 S100 calcium binding protein A16, transcript variant X3

XM_005677499.3
S100A16 S100 calcium binding protein A16, transcript variant X1

XM_018046137.1
S100A16 S100 calcium binding protein A16, transcript variant X2

Fig. S3. A) The BLAST search interface. B) The results page. Clicking “View in JBrowse” (circled in red on the figure) opens JBrowse in a new browser tab. C) BLAST HSPs showing in JBrowse. If you are already logged into Apollo, the Apollo view, rather than JBrowse, will open.

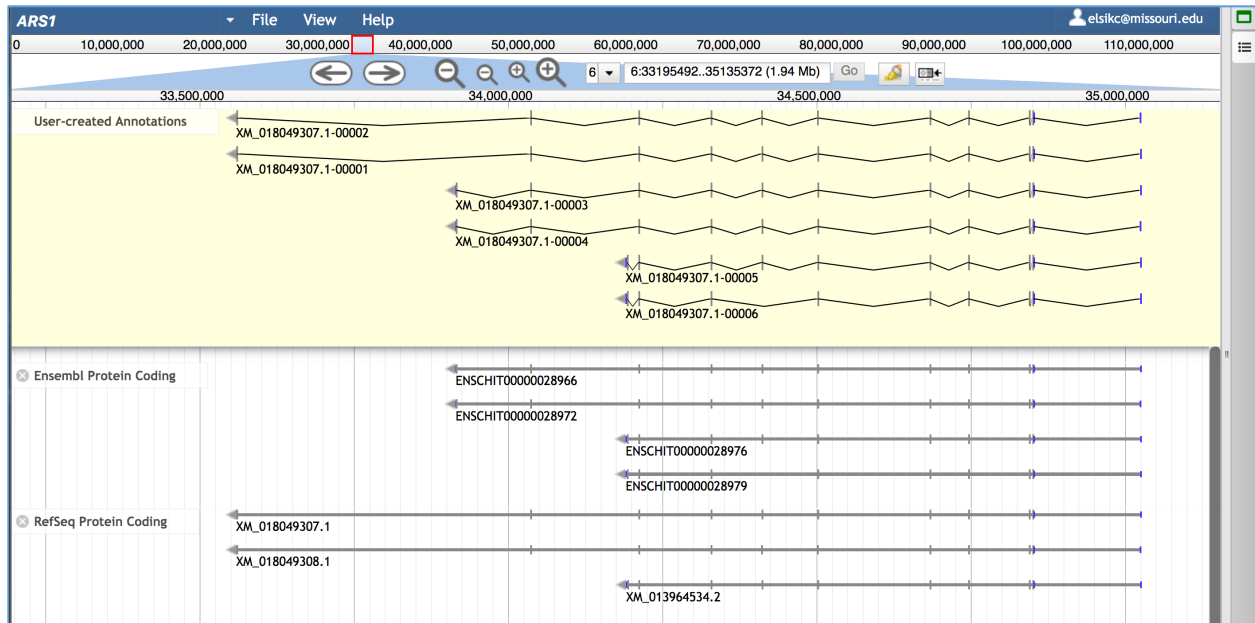


Fig. S4. Transcripts that have been dragged to the Editing Area. XM_018049307.1 was the first transcript to be added, so this identifier is used as the gene name in the Annotation Information Panel and the base name for all the transcripts, including those initiated with Ensembl transcripts.

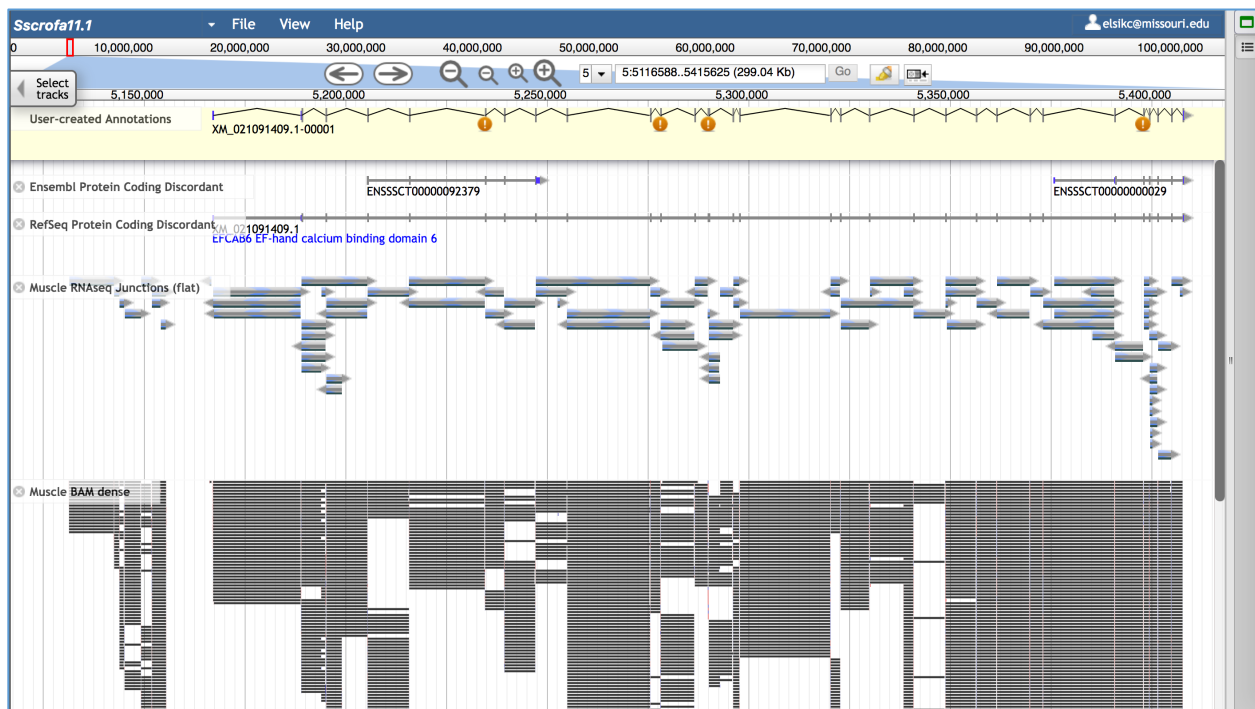


Fig. S5. An annotation with non-canonical splice sites highlighted by exclamation marks. This also shows the tracks that are usually used to verify splice sites (flat RNAseq junctions and dense RNAseq BAM read alignments).

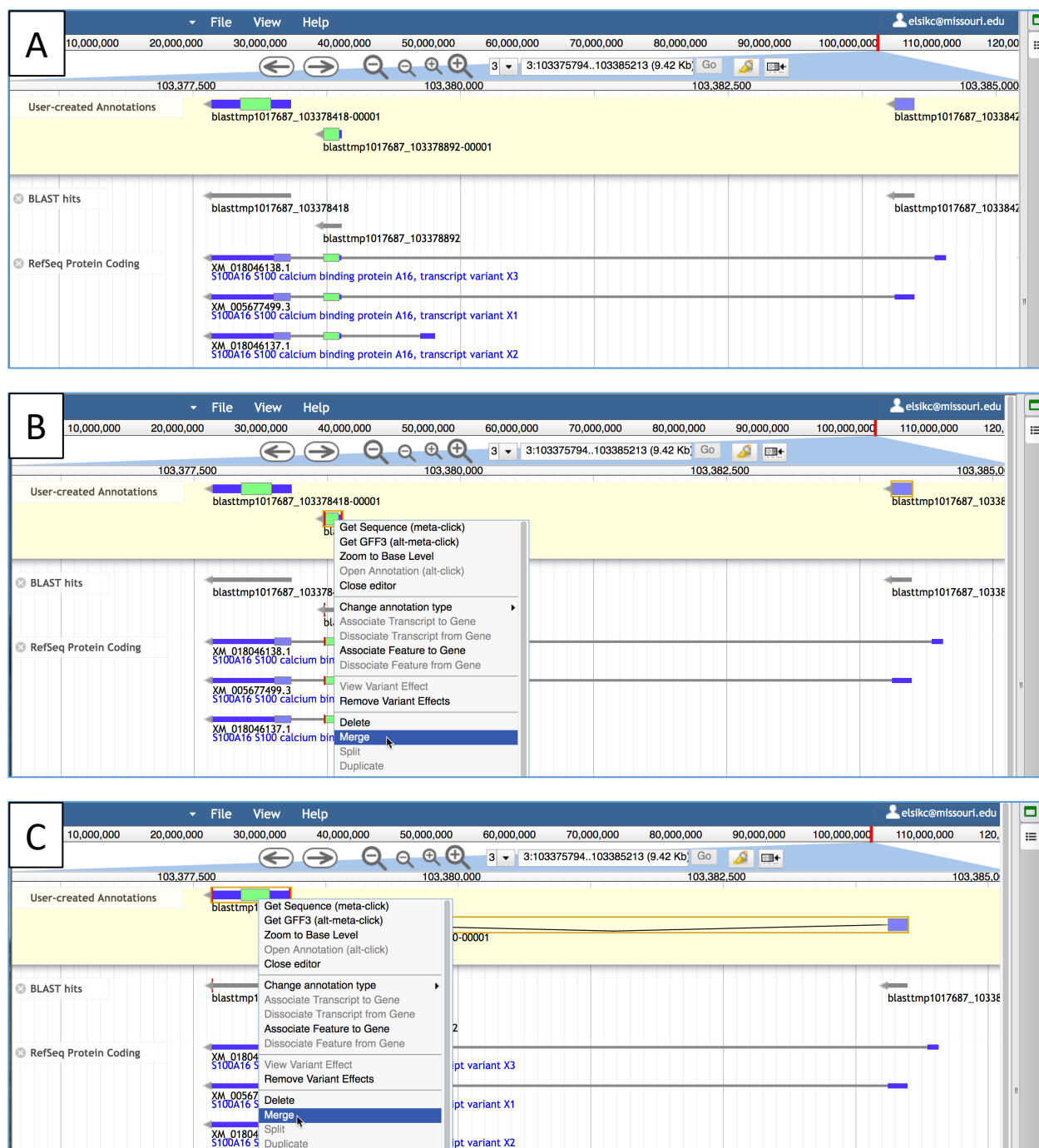


Fig. S6. A) BLAST HSPs have been dragged to the Editing Area to create an annotation, following from the BLAST search shown in Fig. S3. B) After clicking the 5' exon (shown surrounded by a gold box on the far right), the middle exon is right-clicked and "Merge" is selected. C) After the two are merged in panel B, the same procedure is used to merge those to the 3' exon (on the far left).

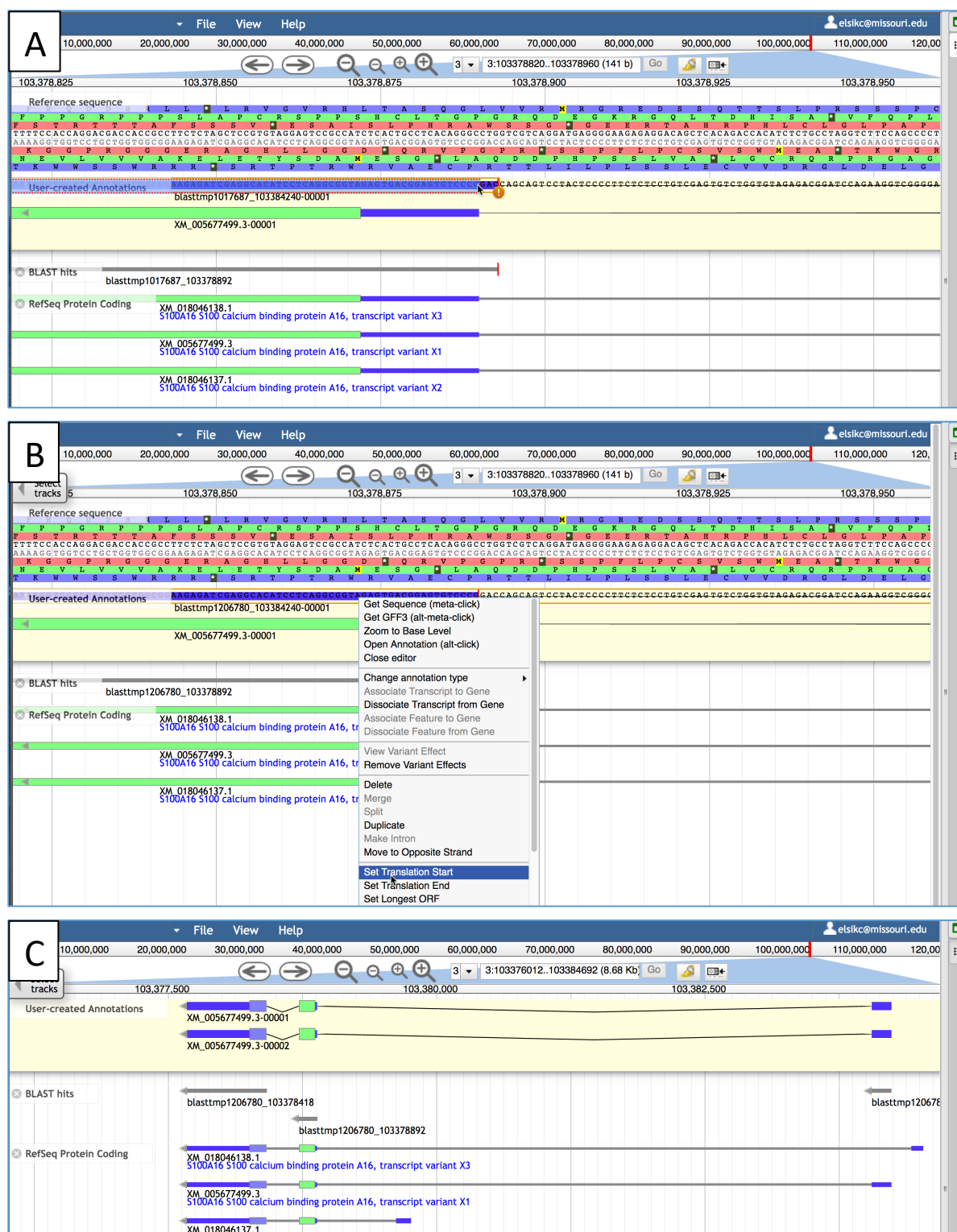


Fig. S8. Corrections to the annotation created by merging BLAST HSPs. A) The edge in the middle exon is corrected by dragging it to the left. B) The translation start is reset. The last correction, not shown, is a correction to the left edge of the 5' exon. C) The corrected annotation. Once the corrections resulted in a CDS with a reading frame that matched the corresponding CDS in the NCBI transcript, the base name of the BLAST-based annotation was automatically changed to the NCBI id.