## Annotation Example

This example shows the annotation of STARD9 in *S. scrofa*, for which there is a split/merge disagreement between Ensembl and RefSeq. This problem was first discovered by viewing the Gene Prediction Problems tracks in chromosome 1. To do this example, you must use the *Sus scrofa* demo browser, which you can access using the Tutorial and Demo pulldown menu (Figure. 1).

Home	BLAST	JBrowse/Apollo 👻	Tutorial and Demo 🕶	Apollo Registration -	How to Cite	Elsik Lab Home				
Wolcor	no to Ag	nimalConomos	About JBrowse/Apollo Track	s						
Welcome to AgAnimalGenome			Apollo Tutorial							
AgAnimalG	enomes.org p	rovides tools based on .	About Apollo Demos	stock genomics rese	stock genomics researchers to browse and manually					
annotate livestock genome assemblies. While a lo data tracks (e.g. genes, gene expression, functio		Bos taurus Demo		new annotations with Apollo, anyone can view the genomes and all						
data tracks pulldown m		jene expression, functio	Bubalus bubalis Demo	) in JBrowse without	logging in, via the	inks in JBrowse/Apollo				
•			Capra hircus Demo							
		ol is a plugin for JBrowse submit gene corrections	Equus caballus Demo		-	es or transcripts. If you				
	0	orm, we will activate you	Gallus gallus Demo	0 1 0	gistration page in the Apollo Registration pulldown menu. After you Once your account is activated, you can use the JBrowse/Apollo					
pulldown m	nenu to select	a genome, then login us	Ovio aries Demo	r of the JBrowse view	v.					
Those who	wish to view a	and export existing user	Sus scrofa Demo	on any of the brows	ers in the JBrowse/	Apollo pulldown menu with				
the userna	me <b>aganimal_</b>	guest and password gu	S. This will not allow earling	whes or annotating ne	w genes.					
Demo brov	vsers are availa	able for practice annotati	on. See About Apollo Demos	under the Tutorial and D	emo pulldown mer	u.				
Ploase con	taat us if you b	nave questions or sugge	tions							
	tact us il you i	lave questions of sugget	50015.							
				1759896 as part of the FAA						

Figure 1. Selecting the Sus scrofa demo browser.

After selecting the browser in the pulldown menu (Figure 1), you will see a JBrowse view and will not yet be logged into Apollo (Figure 2). The first time you access the browser, no tracks will be visible. For future access, the last tracks you viewed in the previous session will be visible.

Click the *Select tracks* button (circled in red in Figure 2) to open the Faceted Track Selector. Once in the Faceted Track Selector, click the Gene Prediction Problems category on the left for filter for those tracks, then check the box to the left of both Gene Prediction Problems tracks (Ensembl Protein Coding Discordant and RefSeq Protein Coding Discordant). Checking the box next to Data Type above the table will cause all the rows to be selected. Selected rows will be highlighted with a blue background. Click *Back to Browser* above the table to close the Faceted Track Selector.

Back in the browser you will see the selected tracks. Most likely you will be zoomed out and the tracks will either look like histograms or many small gray arrows, depending on the gene density in each track (Figure 3A). Enter STARD9 in the search box and click Go (Figure 3B) to navigate to this gene. Searching for this gene automatically opens the RefSeq Protein Coding track, because STARD9 is found in this track as well as the RefSeq Protein Coding Discordant track. The RefSeq Protein Coding track is not needed, so remove it by clicking the X in the track label (Figure 3C).

Apollo File	View Help				Login
0	50,000,000	100,000,000	150,000,000	200,000,000	250,0 🛆
Select		🗲 🏵 ପ୍ରଷ୍ପ୍	1 • 1:1274330532 (274.33 Mb)	Go 🔬 🎫	
tracks	50,000,000	100,000,000	150,000,000	200,000,000	250,000,000

Ay Tracks Currently Active		Back to browser	X Clear All Filte	rs Contair	ns text			1,4	90 tracks	В
Recently Used		Data Type 🔺	Key	BTO Term	Uberon Term	Organ System	SRA Experiment	Biosample	Bioproject	Specimen Tag
Data Type 27 ATAC-seq 16 ChIP-seq (CTCF)		ATAC-seq	ATAC Adipose tissue, SRX9176777	adipose tissue	adipose tissue	adipose tissue	SRX9176777	SAMN16245401	PRJNA665194	adipose_repP348
28 ChIP-seq (H3K27ac) 28 ChIP-seq (H3K27me3) 27 ChIP-seq (H3K4me1) 28 ChIP-seq (H3K4me3)		ATAC-seq	ATAC Adipose tissue, SRX9176778	adipose tissue	adipose tissue	adipose tissue	SRX9176778	SAMN16245400	PRJNA665194	adipose_repP350
2 Chromatin State 6 Gene Prediction 2 Gene Prediction Problems	5	ATAC-seq	ATAC Cerebellum, SRX9176780	cerebellum	cerebellum	nervous	SRX9176780	SAMN16245398	PRJNA665194	cerebellum_repP 348
7 RNAseq Combined Density 16 RNAseq PE Daw dense 16 RNAseq PE BAM draggable 16 RNAseq PE Junctions (arcs)		ATAC-seq	ATAC Cerebellum, SRX9176781	cerebellum	cerebellum	nervous	SRX9176781	SAMN16245397	PRJNA665194	cerebellum_repP 350
16 RNAseq PE Junctions (flat) 16 RNAseq PE StringTie 16 RNAseq PE StringTie 16 RNAseq PE XYPlot		ATAC-seq	ATAC Cerebral cortex, SRX9176784	cerebral cortex	cerebral cortex	nervous	SRX9176784	SAMN16245394	PRJNA665194	cerebral_cortex_ repP348
1 Repeats 2 Variation Organ System		ATAC-seq	ATAC Cerebral cortex, SRX9176785	cerebral cortex	cerebral cortex	nervous	SRX9176785	SAMN16245393	PRJNA665194	cerebral_cortex_ repP350
18 NA 12 abdomen 92 adipose tissue		ATAC-seq		cytotoxic T-lymphocyte	NA	hematopoietic, immune	ERX2628506	SAMEA1088323	PRJEB27111	SAMEA1088323
66 cardiovascular 84 digestive 28 digestive, hematopoietic, immune		ATAC-seq		cytotoxic T-lymphocyte	NA	hematopoietic, immune	ERX2628507	SAMEA1088337	PRJEB27111	SAMEA1088337
<ul> <li>24 endocrine, hematopoietic, immune</li> <li>50 female reproductive</li> </ul>		ATAC-seq		cytotoxic T-lymphocyte	NA	hematopoietic, immune	ERX2628508	SAMEA1088343	PRJEB27111	SAMEA1088343
6 hematopoietic, immune 6 lymphatic 42 male reproductive		ATAC-seq		cytotoxic T-lymphocyte	NA	hematopoietic, immune	ERX2628509	SAMEA1088362	PRJEB27111	SAMEA1088362
188 muscular 276 nervous 92 respiratory 66 urinary		ATAC-seq		helper t-lymphocyte	NA	hematopoietic, immune	ERX2628502	SAMEA1088372	PRJEB27111	SAMEA1088372

Select Tracks											)	Login
My Tracks     Currently Active		Back to browser	X Clear All Filt	ers Contai	ins text				2 matching tracks		•	
Recently Used	1	Data Type	Кеу	BTO Term	Uberon Term	Organ System	SRA Experiment	Biosample	Bioproject	Specimen Tag	3	Sele trac
Data Type         X           27 ATAC-seq         TAC-seq           16 ChIP-seq (CTCF)         28 ChIP-seq (H3K27ac)           28 ChIP-seq (H3X27me3)         27 ChIP-seq (H3X47me1)           28 ChIP-seq (H3K4me1)         26 ChIP-seq (H3K4me3)	✓ ✓	Gene Prediction Problems Gene Prediction Problems	Ensembl Protein Coding Discordant RefSeq Protein Coding Discordant	NA	NA	NA	NA NA	NA	NA	NA NA		
22 Chromatin State 6 Gene Prediction 2 Gene Prediction Problems 7 RNAseq Combined Density 216 RNAseq PE BAM draggable 216 RNAseq PE BAM draggable												

Figure 2. A) The initial JBrowse view showing no tracks. The Faceted Track Selector is opened by clicking *Select tracks* on the left of the browser (circled in red). B) The Faceted Track Selector. Clicking Gene Prediction Problems on the left (circled in red) to filter for those tracks. C) The Faceted Track Selector after filtering for Gene Prediction Problem tracks. Both tracks (Ensembl Protein Coding Discordant and RefSeq Protein Coding Discordant) are selected. Once a track is selected, its row in the table is highlighted in blue. The Faceted Track Selector is closed by clicking *Back to browser* (circled in red).

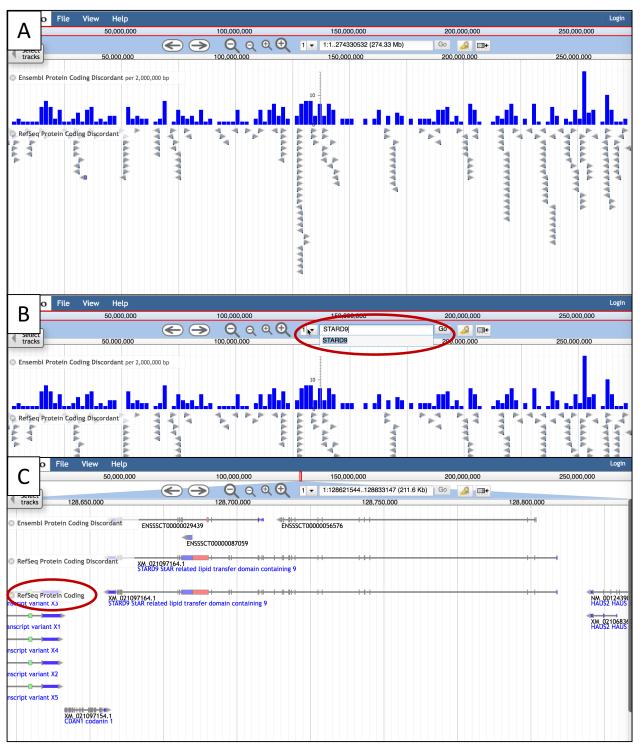


Figure 3. A) The view after selecting the Gene Prediction problem tracks. When you are zoomed out the tracks with a high density of features, similar to the Ensembl Protein Coding Discordant track, look like histograms representing gene density, similar to Ensembl Protein Coding Discordant track. Tracks with lower density, such as the RefSeq Protein Coding Discordant track show individual genes appearing as gray arrows. B) Entering a gene symbol (STARD9) in the search box (circled in red) to navigate to that gene. C) The result of navigating to the STARD9 gene. The RefSeq Protein Coding Gene track was opened automatically because STARD9 is also found in that track. The track is not needed and can be removed by clicking the X in the track label (circled in red).

Figure 4A shows the gene region of interest. Verify that the Ensembl transcripts are considered two different genes by viewing the details for each one. For each transcript, right-click the transcript and then select *View details*. Figures 4B and 4C show that the transcripts are from two different genes. This is an example of a split/merge disagreement between Ensembl and RefSeq and should be annotated with Apollo. Click Login at the upper right of the browser window to login to Apollo (Figure 4A). The credentials for the Sus scrofa demo browser are username: aganimal\_demo\_guest and password: aganimal\_demo.

<b>^</b>	o File View Help					Login
A	50,000,000	100,000,000		150,000,000	200,000,000	250,000,000
Select	1		<b>£</b> 1.	1:1286060661288196		
tracks	128,650,000	128,	700,000		128,750,000	128,800,000
🙁 Ensemb	ol Protein Coding Discordant	ENSSSCT00000029439	087059	ENSSSCT00000056576	5	
RefSeq	Protein Coding Discordant	XM 02:1097164.1 STARD9 SLAR related lipid transfer domain	n containing 9			
mRNA ENS	SSCT00000029439		×	mRNA ENSSSCT0000	0056576	ĸ
Primary	Data		B	Primary Data		C
Name	ENSSSCT0000029439	l		Name	ENSSSCT00000056576	
Туре	mRNA			Туре	mRNA	
Posit	ion 1:128657567128708952 (	(- strand)		Position	1:128715231128801764 (- strand)	
Leng	th 51,386 bp			Length	86,534 bp	
Attribute	25			Attributes		
Gene_	_id ENSSSCG0000022039			Gene_id	ENSSSCG00000044127	
Id	ENSSSCT00000029439			Id	ENSSSCT00000056576	
Seq_i	<b>d</b> 1			Seq_id	1	
Source	e Ensembl			Source	Ensembl	
Transo	cript_biotype protein_coding			Transcript_bio	type protein_coding	
Transo	cript_id ENSSSCT00000029439			Transcript_id	ENSSSCT00000056576	
Versio	<b>4</b>			Version	2	
Regio	n sequence		_	Region sequen	ce	
AG GC GC	SGAAGGATCGATTCGGATCAGGAACAG CTTGTGGCAGGTACTCATGCCTGATG SAAGGAATCCTCAAGGTCTGAGGGGTT TGATCGGAGGAGGAGCCACCCTGTGGA	strand) class=mRNA length=513 SACATTGGTGAGTGACAGAGCGTACTGGTC GGTGTCTGCAGAGGAACCTCGTGGTGCCCT TGTCTGCTTTTCCTGCCAGCCTATTGGTGGTCT TGACTGGACTTCAGGTGGTTTCAGGTGGT TTGACTGGACTTCAGGTGGTTTCAGGTGGT	TGAC TGGA GGGA GGCT	ATGTCTGTC AGGGATTGA AACTCCCCA AAATGAGTC	715231128801764 (- strand) c sacctacAccAccAcctactoracgeogram Acctecarcorcacgertactartactacata Gortorcettectactactactactacata Gortorcettectactactactacgeotte company and accessing and accessing a strange	ATCCTTAACTCACAGAGTGAGGCC TCGTTTCTGCTGTGCCATGATGGG ACAGGAGTTCCTGTTGTTCAGTGG ATTCCTGGCCTCATTCAGTGAATT

Figure 4. A) JBrowse view after removing the RefSeq Protein Coding track. The login button is circled in red on figure right. B) The information panel shown when right-clicking the Ensembl transcript on the right (ENSSSCT00000029439), showing the gene id ENSSSCG0000022039. C) The information panel shown when right-clicking the Ensembl transcript on the right (ENSSSCT00000056756), showing the gene id ENSSSCG0000005676, confirming that these transcripts are considered to be encoded by different genes.

After logging into to Apollo, the view is split between the browser on the left and the Information panel on the right. Also the Select tracks button is no longer visible (Figure 5A). Click the small list icon in the upper left of the Information Panel to bring back the Select tabs button (Figure 5A). Although the Information Panel includes a Tracks tab that lists the tracks, this list is not easily viewed or search. It is recommended that you use the Faceted Track Selector rather than the Tracks tab in the Information Panel.

After bringing back the Select tracks tab, click the red X in the upper left of the Information Panel to close the panel (Figure 5B).

	11.1 - File View Help 🕹 elsikc@missouri.edu	Sus scrofa (Sscrofa11.1)	~	4	elsikc@missour	•
IA	50,000,000 100,000 150,000 200,000 250,000,000	Annotations Tracks Ref Seque	ence Search			
	Q Q Q Q Q Q 1 ▼ 1:128624835128838442 (213.61 Kb) Go Ø 0 0 □ €	Show JBrowse Track Selector Show All Show Visibl	le Only			
5,000	128,750,000	Annotation Name	D ID All Types	GO	т	
User-	created Annotations	Amotator Name	Air Types		GP	
Enser	nbl Protein Coding Discordant,	1	=	Prov		
	ENSSSCT00000087059	Rows 25 ~ 📧 🔍	1-22 of 22	() ()		
	ENSSSCT0000056576	Name	Seq	Type Length	<ul> <li>Updated</li> </ul>	
	EN555C10000036376	XM_021071754.1	1	gene 10,066	Apr 23, 2023	•
RefSe	q Protein Coding Discordant 021097164 1	XM_021091409.1	5	gene 241,137	Apr 23, 2023	
	q Protein Coding Discordant 021097164.1 STARD9 StAR related lipid transfer domain containing 9	XM_005655372.3	4	gene 345,685	Apr 23, 2023	
		XM_021086527.1	3	gene 117,461	Apr 23, 2023	•
		NM_001146130.2	8	gene 38,689	Apr 23, 2023	•
		XM_013994153.2	14	gene 55,970	Apr 21, 2023	•
		XM_021073703.1	14	gene 77,787	Apr 21, 2023	•
		ENSSSCT0000060735	14	gene 92,972	Apr 21, 2023	
						1
		VM_042092924.2	11	00.000	Apr 24 2022	
	11.1 - File View Help 🚣 elsikc@missouri.edu	VM 012002921 2 Sus scrofa (Sscrofa11.1)	1 <i>A</i>		elsikc@missour	
L R	11.1         ▼ File         View         Help         ▲ elsikc@missouri.edu           50,000,000         100,000,000         150,000,000         200,000,000         250,000,000		ence Search		ABE // 01/12	
В	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000	Sus scrofa (Sscrofa11.1)	• ence Search		ABE // 01/12	
B	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000	Sus scrofa (Sscrofa11.1)			ABE // 01/12	
	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000	Sus scrofa (Sscrofa 11.1) Annotations Tracks Ref Seque Show All Show Visible			elsikc@missour	
track	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 →	Sus scrofa (Sscrofa11.1) Hannotations Tracks Ref Seque Show All Show Visible			elsikc@missour	
track	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000	Sus scrofa (Sscrofa 11.1) Annotations Tracks Ref Seque Show All Show Visible			elsikc@missour	
track	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000	Sus scrofa (Sscrofa 11.1) Annotations Tracks Ref Seque Show All Show Visible			elsikc@missour	
track	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 Q Q Q Q 1 v 1:128624841128838449 (213.61 Kb) Go S Totological control of the second seco	Sus scrofa (Sscrofa 11.1) Annotations Tracks Ref Seque Show All Show Visible			elsikc@missour	
track	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 00 000,000 1:128624841128838449 (213.61 Kb) Go C C C C C C C C C C C C C C C C C C C	Sus scrofa (Sscrofa11.1)         Annotations       Tracks       Ref Seque         Show All       Show Visible         Annotation Name       Rows       25 ✓       🗶         Name       Name       Name       Name	IE Only All Types	GO Prov	elsikc@missour	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Image: Show All Show Visible         Annotations Name         Rows 25 ~         Name         XM_021071754.1	Ie Only ID All Types 1-22 of 22 Seq 1	GO Prov Type Lengtt gene 10,066	T GP Updated Apr 23, 2023	• • • • • • • • • • • • • • • • • • •
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Image: Show All Show Visible         Annotations         Tracks         Ref Seque         Annotation Name         Rows       25 ~         Name         XM_021071754.1         XM_021091409.1	le Only ID All Types 1-22 of 22 Seq 1 5	GO     Frov     Frov	v       userial         v       userial         gp       userial         v       updated         Apr 23, 2023       Apr 23, 2023	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 150,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 100,000 100,000 50,000 000 100,000 100,000 50,000 000 100,000 100,000 50,000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Annotations       Tracks       Ref Seque         Show All       Show Visibil         Annotation Name         Rows       25 •       Image: Compare the second sec	le Only ID All Types 1-22 of 22 Seq 1 5 4	GO Prov Type Lengtt gene 10,066 gene 241,137 gene 345,685	T           GP           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Annotations       Tracks       Ref Seque         Show All       Show Visible         Annotation Name       Rows       25 •         Name       XM_021071754.1       XM_021091409.1         XM_0026655372.3       XM_021086527.1	le Only D D All Types 1-22 of 22 Seq 1 5 4 3	GO Prov Type Lengtr gene 10,066 gene 241,137 gene 345,685 gene 117,461	T           GP           Vupdated           Apr 23, 2023	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Image: Show All Show Visible         Annotations         Tracks       Ref Seque         Show All       Show Visible         Annotation Name         Rows       25 ~         Name         XM_0210911754.1         XM_021091409.1         XM_02109165527.3         XM_021086527.1         NM_001146130.2	le Only ID All Types 1-22 of 22 Seq 1 5 4 3 8	GO           Prov           Prov           gene         10,066           gene         241,137           gene         245,685           gene         117,461           gene         117,461           gene         38,689	elsikc@missour           GP           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023           Apr 32, 2023           Apr 32, 2023           Apr 32, 2023           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Annotations       Tracks       Ref Seque         Show All       Show Visibil         Annotation Name       Annotation Name         Rows       25 v       Image: Comparison of the strength of the strengt of the strength of the strength of the strength of the	le Only ID All Types 1-22 of 22 Seq 1 5 4 3 8 14	v         GO           Prov         Prov           Type         Length           gene         10,066           gene         241,137           gene         345,685           gene         345,685           gene         35,689           gene         55,970	• elsikc@missour           • elsikc@missour           • Updated           Apr 23, 2023           Apr 21, 2023	
User-	50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 50,000,000 100,000,000 150,000,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 150,000 200,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000,000 100,000 250,000,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000,000 100,000 100,000 50,000 000 000 000 000 50,000 000 000 000 000 50,000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 50,000 000 000 000 000 000 000 000 000 0	Sus scrofa (Sscrofa11.1)         Image: Show All Show Visible         Annotations         Tracks       Ref Seque         Show All       Show Visible         Annotation Name         Rows       25 ~         Name         XM_0210911754.1         XM_021091409.1         XM_02109165527.3         XM_021086527.1         NM_001146130.2	le Only ID All Types 1-22 of 22 Seq 1 5 4 3 8	GO           Prov           Prov           gene         10,066           gene         241,137           gene         245,685           gene         117,461           gene         117,461           gene         38,689	elsikc@missour           GP           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023           Apr 32, 2023           Apr 32, 2023           Apr 32, 2023           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023           Apr 23, 2023	

Figure 5. Apollo view after logging in. The view is split between browser on the left and the Information Panel on the right. Information Panel has tabs for Annotations, Tracks, Ref Sequence (chromosome ids) and BLAT Search. The JBrowse Select tracks button that provides access to the Faceted Track Selector can be toggled on by clicking the list icon right below the red X (circled in red on the figure). B) The Information Panel can be hidden by clicking the red X in the upper left of the panel (circled in red on the figure). When the Information Panel is closed, the red X is replaced by a green-bordered square icon that can be used to re-open the Information Panel (not shown).

Once you have closed the Information Panel, you can zoom further into the gene region as shown in Figure 6.

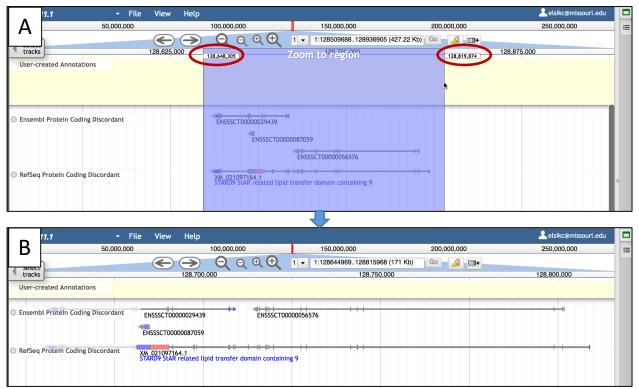


Figure 6. A) The browser after closing the Information Panel. The view is split vertically between the Editing Area with a light yellow background on the top and the Evidence Area with a white background where JBrowse tracks are viewed on the bottom. We can zoom to the region of interest by placing the cursor in the lower number line and dragging it across the region of interest. While doing so, the coordinates of the zoomed view show just below the number line (circled in red on the figure). B) The zoomed-in view.

The next step is to determine which RNAseq experiments will be useful in resolving the split merge disagreement. First, we will use one or more RNAseq Combined Density tracks to identify RNAseq experiments to view. Use the *Select tracks* button to open the Faceted Track Selector. Click RNAseq Combined Density on the left to filter for those tracks (Figure 7A). If the Gene Prediction Problem category is still highlighted, you can click it to un-highlight it. Doing so will remove those tracks from the Faceted Track Selector table, but will not remove them from the browser. The RNAseq Combined Density Tracks must be viewed one-at-a-time because they are large and computationally intensive. We select the track at the top of the list to view first, and then click *Back to browser*. Figure 7B shows the selected track which looks like heatmaps, with darker blue indicating higher read density. On the left side each heatmap is a green rectangle that can be moused over to view the tissue and experiment accession (Figure 7B). We are particularly interested in experiments showing darker blue matching the exons in question (those present in RefSeq but not Ensembl).

After taking note of an informative RNAseq experiment as well as the Bioproject of the combined density track, close the RNAseq Combined Density track. If there were no informative experiments in this track, we would select the next combined density track in the Faceted Track Selector table to view. However, we did find informative experiments, so we will not look at another combined density track. Use the Select tracks tab to open the Faceted Track Selector. Un-highlight *RNAseq Combined Density* and

then click on the Data Type RNAseq PE Junctions (arcs) and the Bioproject PRJEB14330. Enter "lung" in the search box and select both tracks (Figure 8A), then close the Faceted Track Selector. In the browser, zoom in to the discordant region using the small plus sign above the browser (Figure 8B). The arcs appear to be connecting the exons in question (Figure 8C).

t Tracks										Help	uri.edu
A ks ently Active		Back to browser	🗙 Clear All Fil	ters Conta	ins text				7 matching tra	cks	
Recently Used		Data Type 🔺	Кеу	BTO Term	Uberon Term	Organ System	SRA Experiment	Biosample	Bioproject	Specimen Tag	Select track
Data Type     X       27     ATAC-seq       16     ChIP-seq (CTCF)       28     ChIP-seq (H3K27ac)	✓	RNAseq Combined Density	RNAseq Combined Density PRJEB14330	NA	NA	NA	NA	NA	PRJEB14330	NA	
<ol> <li>ChIP-seq (H3K27me3)</li> <li>ChIP-seq (H3K4me1)</li> <li>ChIP-seq (H3K4me3)</li> <li>Chromatin State</li> <li>Gene Prediction</li> </ol>		RNAseq Combined Density	RNAseq Combined Density PRJEB19268	NA	NA	NA	NA	NA	PRJEB19268	NA	
2 Cone Prediction 7 RNAseq Combined Density 216 RNAseq PE BAM dense 216 RNAseq TE Dawn draggable		RNAseq Combined Density	RNAseq Combined Density PRJEB19386	NA	NA	NA	NA	NA	PRJEB19386	NA	
216       RNAseq PE Junctions (arcs)         216       RNAseq PE Junctions (flat)         216       RNAseq PE StringTie         216       RNAseq PE StringTie         216       RNAseq PE XYPlot		RNAseq Combined Density	RNAseq Combined Density PRJEB27455	NA	NA	NA	NA	NA	PRJEB27455	NA	
1 Repeats 2 Variation • Organ System 7 NA		RNAseq Combined Density	RNAseq Combined Density PRJEB37735	NA	NA	NA	NA	NA	PRJEB37735	NA	
<ul> <li>Bioproject</li> <li>1 PRJEB14330</li> <li>1 PRJEB19268</li> <li>1 PRJEB19386</li> </ul>		RNAseq Combined Density	RNAseq Combined Density PRJNA311523	NA	NA	NA	NA	NA	PRJNA311523	NA	
1 PRJEB27455 1 PRJEB37735 1 PRJNA311523 1 PRJNA665193		RNAseq Combined Density	RNAseq Combined Density PRJNA665193	NA	NA	NA	NA	NA	PRJNA665193	NA	

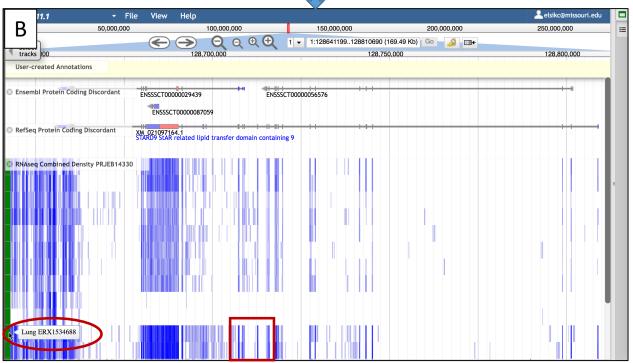
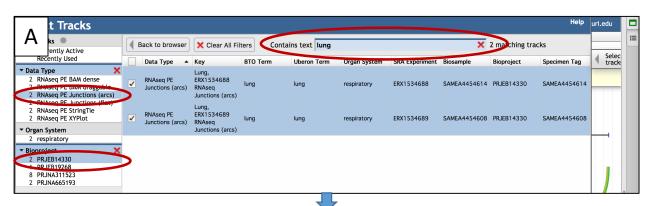
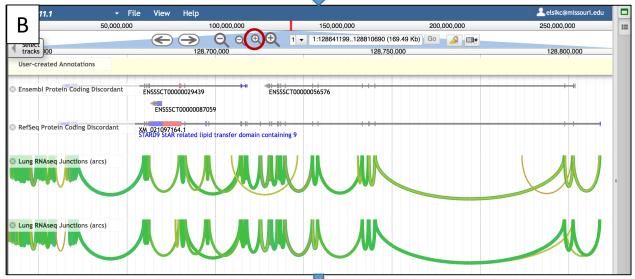


Figure 7. A) The Faceted Track Selector showing the RNAseq Combined Density category selected on the left and one track selected in the table on the right. The combined density tracks are large tracks, and only one can be viewed at a time. B) The highlighted experiment in the track, Lung ERX1534688, and the experiment below it both look informative because there are dark blue marks corresponding to exons that are present in the RefSeq transcript, but not Ensembl (shown in red square).





11.1	<del>-</del> File View Help				🚨 elsikc@missouri.edu 🗖
50,00	0,000	100,000,000	150,000,000	200,000,000	250,000,000
		$\Theta \ominus \Theta \Theta$	1 - 1:128705221128727920 (22.7 Kb)	Go 💋 🎉	
tracks 128,705,000	00	128,710,000	128,715,000	128,720,000	12
User-created Annotations					
S Ensembl Protein Coding Discorda	Int ENSSSCT00000029439		ENSSSCT0000005	5576	
💿 RefSeq Protein Coding Discordar	t XM 021097164.1 STARD9 StAR related	lipid transfer domain containir	g 9		
Lung RNAseq Junctions (arcs)					
Lung RNAseq Junctions (arcs)				M	

Figure 8. A) Selecting RNAseq Junctions (arcs) tracks for Bioproject PRJEB14330, filtered for by entering "lung" in the search box (circled in red). B) JBrowse view showing the arc junction tracks. The thickness and color of the arcs represents the numbers of reads supporting the junction, with lighter green/yellow colors supported by fewer reads and darker green supported by a larger number of reads. The light thin arcs in the upper track may be spurious alignments. C) JBrowse view zoomed in more to the exon that is present in RefSeq but not Ensembl.

The arc junction tracks show support for the RefSeq transcript. For an alternative view, we will look at the same experiments in the dense RNAseq BAM format. Remove the arcs tracks by clicking the X in the track labels. Open the Faceted Track Selector. Un-highlight RNAseq PE Junctions (arcs) and highlight RNAseq PE BAM dense. Enter "lung" in the search box if it is not already entered (Figure 9A). After closing the Faceted Track Selector you will see the dense BAM tracks (Figure 9B). If you are zoomed-out too far, you will get a "Too much data to show" error. In that case, zoom in until the error goes away.

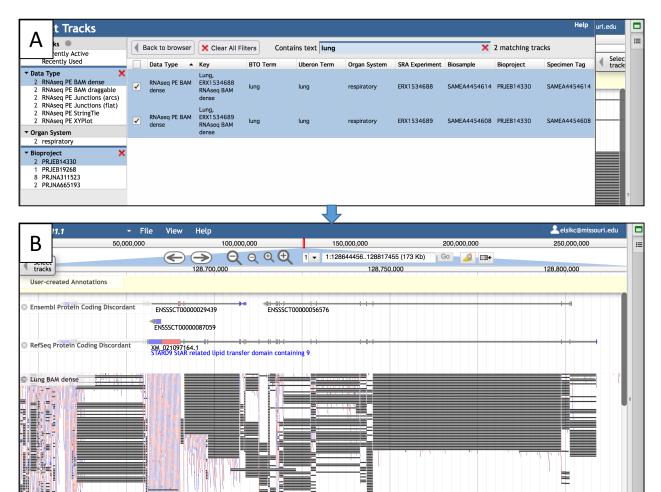


Figure 9. A) Selecting RNAseq BAM (dense) tracks for Bioproject PRJEB14330, filtered for by entering "lung" in the search box. B) JBrowse view showing the dense BAM tracks. The small red and blue rectangles are the aligned portion of the reads. Splice junctions are shown as dark gray lines.

It is easier to focus on splice junctions if you remove unspliced reads from dense BAM tracks. Do so by clicking the downward arrow that becomes visible when you mouse over a track label, and select Hide unspliced reads (Figure 10)

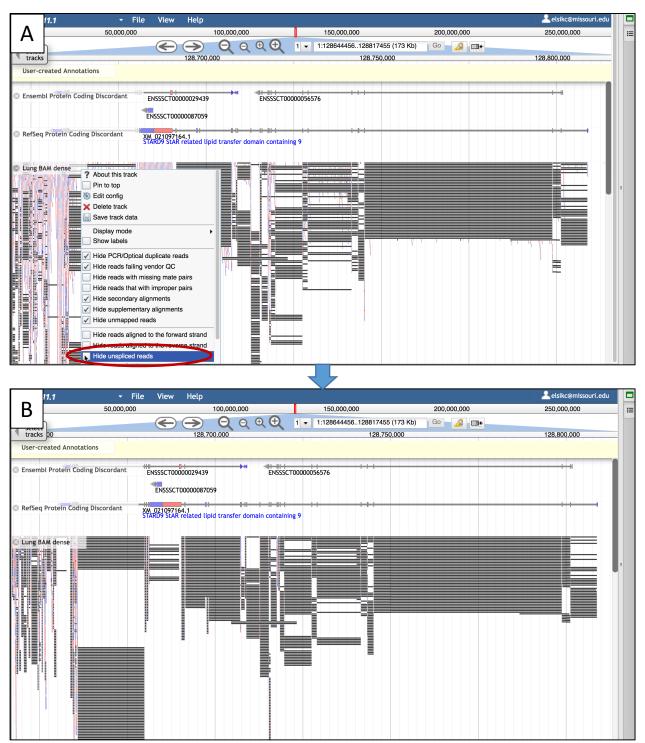


Figure 10. A) Hiding unspliced read alignments in the dense BAM track by clicking the track label and selecting Hide unspliced reads in the pulldown menu (circled in red). B) JBrowse view after hiding unspliced reads.

Zoom in further to investigate the exons and introns present in the RefSeq but not Ensembl (Figure 11). The dense BAM track provides strong support for the RefSeq transcript.

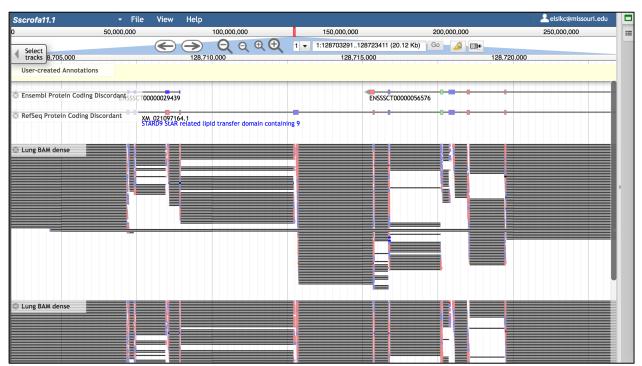


Figure 11. A zoomed-in view of dense RNAseq BAM tracks with unspliced reads hidden. These tracks and the previously viewed arc junctions tracks support the RefSeq gene model as being correct.

Based on the RNAseq evidence, we decide to use the RefSeq transcript to initialize the annotation. Click the RefSeq transcript and drag it up to the Editing Area (Figure 12). The transcript in the editing area is now called an *annotation*.

Sscrofa11.1 -	File View I	Help			Lelsikc@missouri.edu
0 50,000,0	00	100,000,000	150,000,000	200,000,000	250,000,000
Select     Select User-created Annotations     XM_	021097464.1-00001		1   1:128649798128813797 128,75		
S Ensembl Protein Coding Discordant	ENSSSCT00000		CT00000056576		
RefSeq Protein Coding Discordant	XM 021097164. STARD9 StAR rel	ated lipid transfer domain cont	aining 9		

Figure 12. The RefSeq transcript is clicked, then dragged to the Editing Area. Clicking the transcript causes it to be surrounded by a red box and causes the exon edges to be highlighted in red. The edges of exons in other transcripts that match the RefSeq exon edges are also highlighted in red. Notice that the id of the annotation is the id of the original transcript with an extension added. The exclamation mark (seen near the id) indicates a non-canonical splice site.

Once the transcript has been dragged to the editing area, we wish to confirm that the junctions in the RNAseq data perfectly match the introns in the annotation. We will use flat RNAseq junctions to do so. Remove the dense BAM tracks from the browser by clicking the X in the track labels. Open the faceted track selector. Un-highlight *RNAseq PE BAM dense* and highlight *RNAseq PE Junctions (flat)*. Make sure that "lung" is entered in the search box. Select both tracks and close the Faceted Track Selector (Figure 13A) to return to the browser (Figure 13B). Zoom in to the first intron of interest (Figure 13C).

Identify a junction that appears to correspond with the intron in the annotation (Figure 13C). Right click the selected junction to view details. Note the *Score*, which is the number of reads supporting the junction. Also note the start and end coordinates of the junction (Figure 14A). Determine whether the exon boundaries in the annotation is perfectly abutted with junction by placing the cursor in the lower number line and moving it until the red line perfectly aligns with the exon edge, and note the number shown in the small bow next to the cursor (Figure 14B). We see that the exon edge on the left is one less than the junction start coordinate, so it the exon does perfectly abut with the junction. Repeating the same procedure for the exon edge on the right, we see that it also is also perfectly abutted with the junction (Figure 14C). This procedure is repeated for all introns that are present in RefSeq but not Ensembl, and exons that disagree between the two gene sets.

After checking splice junctions, we look for additional issues. Zooming in to the 3' end of the transcript (the far right) we can see that the start codon in the annotation is not in the same location as the original transcript (Figure 15A). Apollo computes start codons based on the largest open reading frame, and does not always reproduce the original coding sequence. Zoom into region as far as possible and click the annotation so you can see the bases in the annotation. The color of the exon in the annotation matches the color of the translated reference sequence above. This transcript is on the minus strand, so we look at the peach colored reading frame below the reference sequence DNA (Figures 15B and 15C).

^	:t Tracks								Help	uri.edu
А	iks 🔍	Back to browser	X Clear All Filters	Contains text	lung		×	2 matching tra	cks	
Re	rently Active Cently Used	Data Type 🔺	Key BTC	Term Uberon	Term Organ System	SRA Experiment	Biosample	Bioproject	Specimen Tag	Selec track
2 RI 2 RI	Type X NAseq PE BAM dense NAseq PE BAM draggable NAseq PE Junctions (arcs) NAseq PE Junctions (flat)	RNAseq PE Junctions (flat)	Lung, ERX1534688 lung RNAseq Junctions (flat)	ling	respiratory	ERX1534688	SAMEA4454614	PRJEB14330	SAMEA4454614	
2 RM 2 RM • Organ	VAseq PE StringTie VAseq PE XYPlot System Ispiratory	RNAseq PE Junctions (flat)	Lung, ERX1534689 lung RNAseq Junctions (flat)	l lung	respiratory	ERX1534689	SAMEA4454608	PRJEB14330	SAMEA4454608	
• Biopro 2 PF 1 PF 8 PF					_					
	1.1	▼ File View	Help	~					alsikc@misso	ouri.edu
В	50,00	00,000	100,000,000		150,000,000		200,000,000		250,000,000	
track	<u></u>		→ Q Q 128,700,000	🕀 🔁 🛛 । •		052 (164 Kb)	ao 🧳 🔝		128,800,00	0
	created Annotations	XM_021097164.1-00001							/	
Ensem	bl Protein Coding Discord			ENSSSCT000000	56576				H	
		ENSSSCT00000	087059							
RefSee	q Protein Coding Discordar	nt XM 021097164. STARD9 StAR re	1 lated lipid transfer do	main containing 9						
Lung	RNAseg Junctions (flat)	JUNC00012757	JUNC00012764	JUNC00012771	JUNC00012784 JUNC000	012786			UNC00012794	
12700	JUNC00012741	JUNC00012758	JUNC00012765	JUNC00012773	JUNC00012785 JU	JNC00012792				
012701	JUNC00012742	JUNC00012760	JUNC000		-	0012787			JUN	C00012795
012702	JUNC00012743	JUNC00012759		00012767 JUNC0001	2782 JUN	00012788			-	C00012798
012703	JUNC00012744	-	-	00012768 JUNC0001		00012790			-	UNC000127
012705	4	-	IC00012763	JUNC00012772	-	C00012789			_	UNC000127
012706	4	-	- 4	C00012769	-	C00012791				
	UNC00012736	301	-	IC00012770	-	JNC00012793				
012703	4		JUN	JUNC00012774						
	4			-						
0012709			Help	JUNC00012775					elsikc@missc	ouri.edu
С	· ·	00,000	100,000,000		150,000,000		200,000,000		250,000,000	
-			🗩 <b>Q</b> વ	⊕ ⊕ 1 -	][	· · · · · · · · · · · · · · · · · · ·	ào 🧳 👰			
track User-c	created Annotations	3,707,500 XM_021097164.1-00001		3,710,000	1	28,712,500		128,715	5.000	
Encorr	abl Protoin Coding Disco-d		•							
	bl Protein Coding Discord	Ensader of the							ENSSSCT000000	056576
RefSe	q Protein Coding Discordar	nt XM 021097164. STARD9 StAR rel	I ated lipid transfer dor	nain containing 9						
Lung f	RNAseq Junctions (flat) 3	JUNC00012768	JUNC00012770						JUNC00012	
	JUNC00012765		JUNC00012769			JUNC	00012771		JU	JNC000127
		JUNC00012766				_				
		JUNC00012767				JUNC	00012773			

Figure 13. A) Selecting RNAseq Junctions (flat) tracks for Bioproject PRJEB14330, filtered for by entering "lung" in the search box. B) JBrowse view showing the flat junctions tracks. C) Zoomed in view showing a junction matching an intron that is present in the RefSeq, but not Ensembl, transcript.

A 50,000,000 100,000,000 150,000,000 200,000	📃 🛃 elsikc@missouri.edu 🗖
	250,000,000
Q Q Q Q Q Q Q 1 ↓ 1:128705251128716600 (11.35 Kb) Go 3	
tracks 128,707,500 128,710,000 128,712,500 128,71	5,000
User-created Annotations XM_02109 JUNC00012770 X	
Primary Data	
Ensembl Protein Coding Discordant ( Name JUNC00012770	ENSSSCT00000056576
	EN333C100000030370
where a rote in county back date	
Position 1:128708970128712699 (- strand)	
C Lung RNAseq Junctions (flat) JUNC	JUNC00012772
JUNC00012765 Attributes	JUNC0001;
Seg id 1	JUNCOUT,
JUNCOL Region sequence	
	JUNC00012774
>1 1:128708970128712699 (- strand) length=3730 GTGAGTGGATGGGCGAACCCCTTCTATCCTTGAGCAGCGGCCCCAGAGGTCAGGGAGGG	
11.1 - File View Help	💄 elsikc@missouri.edu 🗖
B 50,000,000 100,000,000 150,000,000 200,000,000	250,000,000
← → Ϙ Ϙ ⊕ ⊕ 1 <u>• 1:128708945.128709085 (141 b)</u> Go 💋 💷	
tracks 128.708.925 128.708.950 128.709.900	128,709,025
Reference v L F L I R I D P S Y F C C P L L K G K V K O L B S K W T O P	СРОСКУН
V R S V T H O C P V P D P N R S F L F L L S P P K G R S E T T L M R I K M D T A 3 T L C H S P M S C S S E S I L P V P S R E K N N N P N L N K N G H S I 6 TA GEGGAAAA GEGAAAA GEGAAA	A P K G K S '
ICATECCAGACAGTAGATGGACAAAGAATAAGCACGGAGGAAGAATAAAACAACAGGGGAGATAATAATCCTTTTCAAATAATACTCTATTTTTTTT	GACGGGGGTTCCCTTTCAGG!
TRET V V V H O T G S G F P V V V V V V V V V V V V V V V V V V	R G G L S L G
User-created Annotations XM 021097164.1-00001	
Ensembl Protein Coding Discordant ENSSSCT00000029439	
RefSeq Protein Coding Discordant XII. 021097164.1     STARD9 SLAR related lipid transfer domain containing 9	
STARUY SLAK related upto transfer domain containing y	
C Lung RNAseq Junctions (flat)	
Uung RNAseq Junctions (flat) JUNC00012766	
JUNC00012769	
11.1 - File View Help	🚨 elsikc@missouri.edu 🗖
C 50,000,000 100,000,000 150,000,000 200,000	Lelsikc@missouri.edu   □     250,000,000   IE
C 50,000,000 100,000,000 150,000,000 200,000 ← → ⊖ ⊖ ⊖ ⊕ 1 • 1:128712611128712751 (141 b) Go ⊿ □+	250,000,000
C 50,000,000 100,000,000 150,000,000 200,000 C O O O O 1 1 1128712611,128712751 (141 b) Go S I + tracks 128,712,650 128,712,675 1127,720 128,712,726	
C 50,000,000 100,000,000 150,000,000 200,000 C O O O O 1 1 1128712611,128712751 (141 b) Go S I + tracks 128,712,650 128,712,675 1127,720 128,712,726	250,000,000
C         50,000,000         100,000,000         150,000,000         200,000,000           tracks         128,712,650         128,712,675         1         1:128712611.128712751 (141 b)         Go         Go	250,000,000 128,712,750
C         50,000,000         100,000,000         150,000,000         200,000,000           tracks         128,712,650         128,712,675         1         1:128712611.128712751 (141 b)         Go         Go	250,000,000
C       50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       187,12,700       128,712,700       128,712,725         Reference sequence       p       L       <	250,000,000 128,712,750
C         50,000,000         100,000,000         150,000,000         200,000,000           tracks         128,712,650         128,712,675         1         1:128712611.128712751 (141 b)         Go         Go	250,000,000 128,712,750
C       50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       11,712,700       128,712,710       128,712,725         Reference sequence       p       p       p       k       m       s       c       c       m       c       c       m       m       k       s       c       m       m       k       s       c       m       m       k       s       c       m       m       k       m       m       m       k       m       <	250,000,000 128,712,750
C       50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       167,12,700       128,712,725         Reference sequence       p       p       p       c       p       p       c       n       p       p       c       n       n       p       p       c       n       n       p       p       c       n       n       n       p       p       n       n       n       n       p       p       n       n       n       n       p       p       n	250,000,000 128,712,750
50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       11,712,700       128,712,725         Reference sequence       P       P       L       L       K       D       R       G       S       F       T       128,712,750       128,712,725         T       L       L       A       S       P       L       L       K       A       A       Q       G       K       K       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       S       T       F       R       R       S       S       T       F       R       R       S       S       S       R       R       S       S       T       T       R       R       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S<	250,000,000 128,712,750
50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       11,712,700       128,712,755         reference sequence       p       p       p       p       p       p       p       p       128,712,675       127,712,700       128,712,725         reference sequence       p	250,000,000 128,712,750
C       50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       187,712,700       128,712,725         Reference sequence       p       p       p       k       c       p       k       c       p       k       c       n	250,000,000 128,712,750
50,000,000       100,000,000       150,000,000       200,000,000         tracks       128,712,650       128,712,675       11,712,700       128,712,725         Reference sequence       P       P       L       L       K       D       R       G       S       F       T       128,712,750       128,712,725         T       L       L       A       S       P       L       L       K       A       A       Q       G       K       K       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       F       F       R       R       R       S       S       T       F       R       R       S       S       T       F       R       R       S       S       S       R       R       S       S       T       T       R       R       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S<	250,000,000 128,712,750

Figure 14. A) Right-clicking the junction shown in Figure 13 allows you to view details. The *Score* is the number of reads supporting that junction. We can see that the junction starts at coordinate 128,708,970 and ends at 128,712,699. B) A zoomed-in view showing the exon boundary at the left of the junction. Placing the cursor in the lower number line causes a red line to appear. The cursor is positioned so that the red line perfectly matches the exon edge. The small box next to the cursor shows that the last exon coordinate (128,708,969) perfectly abuts the first junction coordinate. C) A zoomed-in view showing the exon boundary at the right of the junction showing that the exon first exon coordinate (128,712,700) perfectly abuts the last junction coordinate. Thus the annotation intron is validated by the RNAseq.

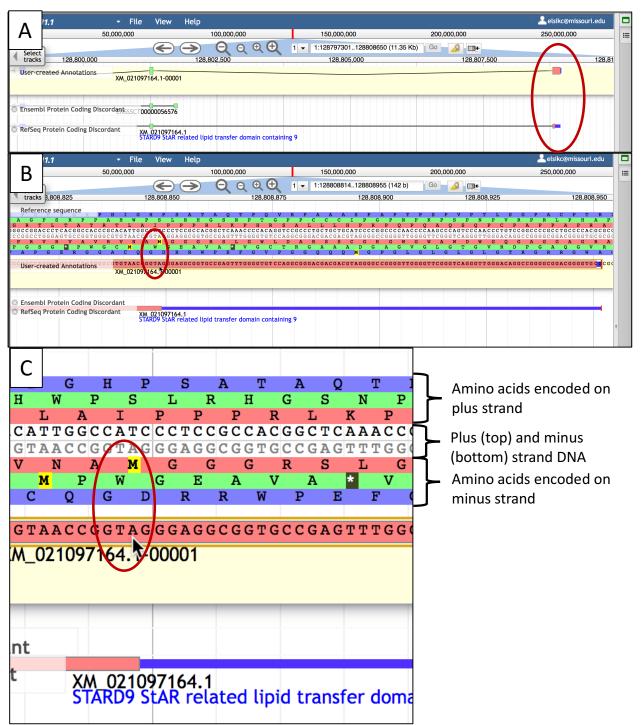


Figure 15. A) Apollo view with the terminal 3' exon in the annotation and original transcript circled in red. Notice that the coding region in the annotation is larger than the coding region in the original transcript. B) Apollo view zoomed-in enough to see the DNA. Clicking on the annotation shows the DNA in it. The Methionine (M) and the start codon (ATG shown in reverse) in the annotation that correspond with the start codon in the original transcript are circled in red. C) The figure is zoomed in to show the cursor pointing to the translation start that corresponds with the original transcript.

The translation start site in the annotation is reset by right-clicking the "A" of the "ATG" in the desired start codon, and then selecting Set Translation Start in the pulldown menu (Figure 16).

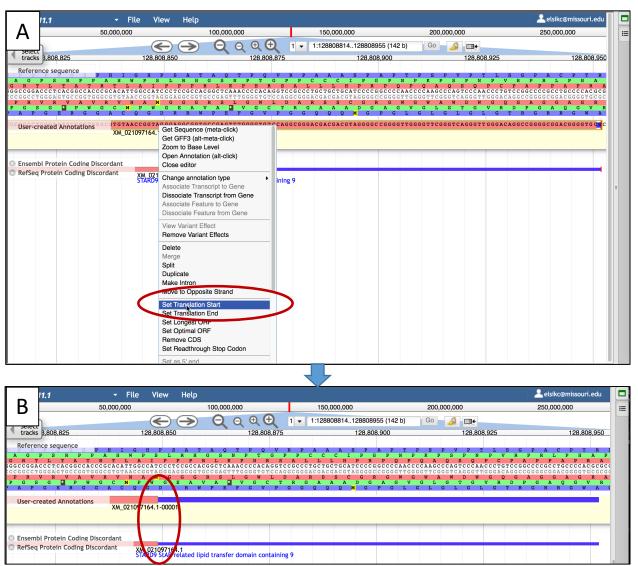


Figure 16. A) Setting the translation start by right-clicking the "A" in "ATG" (shown in reverse), then selecting *Set Translation Start* in the pulldown menu. B) A view showing the corrected translation start site.

The final step is to use the sequence of the annotation in a search of a database of well-known proteins, such as UniProtKB/Swissprot. Obtain the protein sequence of the annotation by right-clicking the annotation and selecting "Get Sequence" in the pulldown menu (Figure 17). Use the NCBI BLAST website (https://blast.ncbi.nlm.nih.gov/Blast.cgi ) and select *Protein BLAST* to search a protein database.

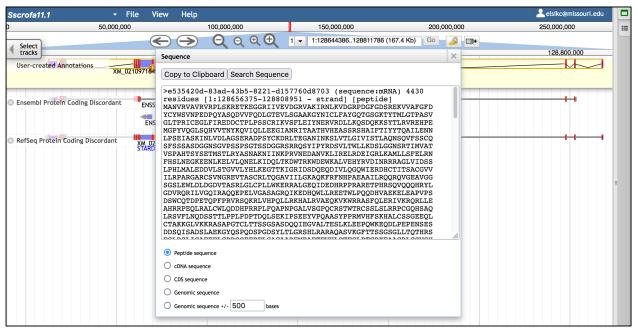


Figure 17. The Sequence panel after right clicking the sequence and selecting Get Sequence in the pulldown menu.