

FlyBase

How High-Throughput
Data are Informing Gene
Models

New high-throughput data

- RNA-Seq junction calls (introns)
- Stranded RNA-Seq coverage data from multiple tissues and cell lines
- Stop-codon readthrough predictions
- A-to-I RNA editing sites
- Transcription start site (TSS) mapping

Integration of High Throughput Data

- Presentations in GBrowse
- Sequence feature reports
- New and improved annotations
- Assessing the data
- Annotation on the genome of transcript features
- Use of gene model comments

GBrowse link on top page

GTGGGCATCCOFTAGATAGCCAAATATTATTATTGTTTCAGATACTCAG
AGGCTTTCAGATGCGY AATTTTTRPAG TGTGGY AAATCAGTGAATTT
ATTCAGGCTTTCAGATGCGY AATTTTTRPAG TGTGGY AAATCAGTGAATTT
AATAATAAATTTTCAGTGCMACTCAGCCGGGGCATCTTCATAGA

FB2012_02, released March 2nd, 2012

FlyBase

A Database of *Drosophila* Genes & Genomes

Home **Tools** Files Species Documents Resources News Help Archives Jump to Gene

 BLAST	 GBrowse >>	 QueryFlyBase	 TermLink	 ImageBrowse	 Batch Download
-----------	----------------	------------------	--------------	-----------------	--------------------

Fast-Track Your Paper

FlyBase Forum

Find a Fly Person

News

- modENCODE webinar MAR 1 | 17 Feb 12
- In Memoriam: Arthur Chovnick | 6 Jan 12
- In Memoriam: James Crow | 6 Jan 12
- FlyBase 101 | 23 Dec 11
- 21 Dros genomes avail | 29 Nov 11
- 26 new species in BLAST | 18 Nov 11
- 2012 Release Schedule | 27 Oct 11
- Nobel Prize: Jules Hoffmann | 6 Oct 11

QuickSearch

Simple Expression Phenotype GO References Data Type

Species: include non-Dmel drosophilid species [QuickSearch help](#)

Search: ID/Symbol/Name All text

Data Class:

°Enter text:

°QuickSearch autocomplete: **Note: Wild cards (*) can be added to your search term**

Commentary [See all commentaries](#)

Community curation effort a big success!

GBrowse: Data Source default is Gene Models/Evidence

- **Click For Ease** Master text styles



FB2012_01, released January 20th, 2012

Home Tools Files Species Documents Resources News Help Archives Jump to Gene Go

Instructions [Help] [Reset] [Bookmark this]

"Landmark or Region Search": sequence region or exact ID symbol (but not full name, for example, ct but not cut), case-sensitive, no wildcards.
"Advanced Search": cytoloactions, symbols (wildcards allowed).
To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: `cnn`, `FBgn0000490`, `X:60000..80000`, `2L:93,000..100,000`, `3L:80,000..100,000`, `3R:80,000..100,000`, `4:20000..50000`.

Search

Switch to Chromosome Map

» Fifth level

Advanced Search: Cytolocation [dropdown] [Start]

Landmark or Region:

2R:9326816..9337978 Search

Report & Analysis tools:

Download Decorated FASTA File [dropdown] Configure... Go

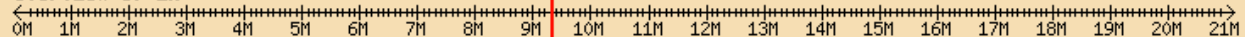
Data Source

'D. melanogaster Gene Models/Evidence' [dropdown]

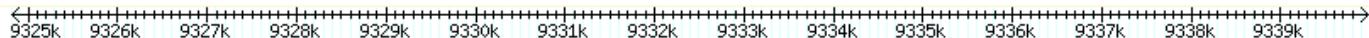
Scroll/Zoom: <<< < - Show 15.16 kb [+ > >>] Flip

Overview

Overview of 2R



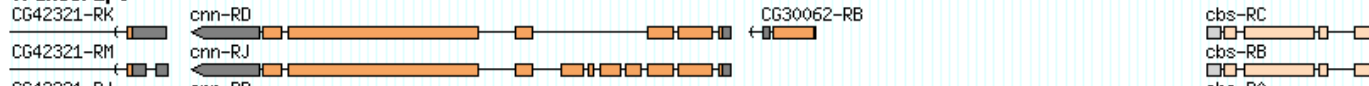
Details



Gene Span



Transcript



GBrowse: Data Sources

Click FlyBase Map

Second level

Third level

Fourth level

Fifth level

Home Tools Files Species Documents Re

Instructions [Help] [Reset] [Bookmark this]
"Landmark or Region Search": sequence region or exact ID symbol (full or not full name, "Advanced Search": cytoloactions, symbols (with IDs allowed).
To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnifi

Examples: [cnn](#), [FBgn0000490](#), [X:60000..80000](#), [2L:93,000..100,000](#), [2R:90,000..100,000](#)

Search

Switch to Chromosome Map >>

Landmark or Region:
2R:9326816..9337978 Search

Data Source
' D. melanogaster Gene Models/Evidence

Overview

Overview of 2R
0M 1M 2M 3M 4M 5M 6M 7M 8M 9M

Details

9325k 9326k 9327k 9328k 9329k 9330k 9331k 9332k

Gene Span
CG42321
cnn
tRNA:CR30509

Transcript
CG42321-RK
cnn-RD
CG42321-RM
cnn-RJ
CG42321-RL
cnn-RL

Landmark or Region:
2R:9326816..9337978 Search

Data Source

- ✓ ' D. melanogaster Gene Models/Evidence
- ' D. melanogaster Expression/Regulation ←
- ' D. melanogaster MegaView
- ' D.melanogaster Gene Disruptions, Stocks and Reagents
- ' D. ananassae
- ' D. erecta
- ' D. grimshawi
- ' D. mojavensis
- ' D. persimilis
- ' D. pseudoobscura
- ' D. sechellia
- ' D. simulans
- ' D. virilis
- ' D. willistoni
- ' D. yakuba
- A. aegypti (mosquito)
- A. gambiae (mosquito)
- A. mellifera (honey bee)
- A. pisum (pea aphid)
- B. mori (silk worm)
- C. elegans (worm)
- C. pipiens (mosquito)
- I. scapularis (deer tick)
- N. vitripennis (wasp)
- P. humanus corporis (human body louse)
- T. castaneum (red flour beetle)



FB2012_01, released January 20th, 2012

Gene Dmel\chinmo

Home Tools Files Species Documents Resources News Help Archives

Jump to Gene

Profile Manager

General Information

Symbol	Dmel\chinmo	Species	<i>D. melanogaster</i>
Name	Chronologically inappropriate morphogenesis	Annotation symbol	CG31666
Feature type	protein_coding_gene	FlyBase ID	FBgn0086758
Gene Model Status	Current	Stock availability	14 publicly available
Also Known As	CG17156		

Genomic Location

Chromosome (arm)	2L	Recombination map	
Cytogenetic map	22A5-22B1	Sequence location	2L:1,651,260..1,704,682 [+]

Genomic Maps

Select View:

modENCODE
[GBrowse](#)



Files

Summary Information

Recent Updates

GBrowse: Expression/Regulation

Click to edit Master text styles

- Section
- Title



gene
extents
annotated
transcripts
natural TE's

cDNAs

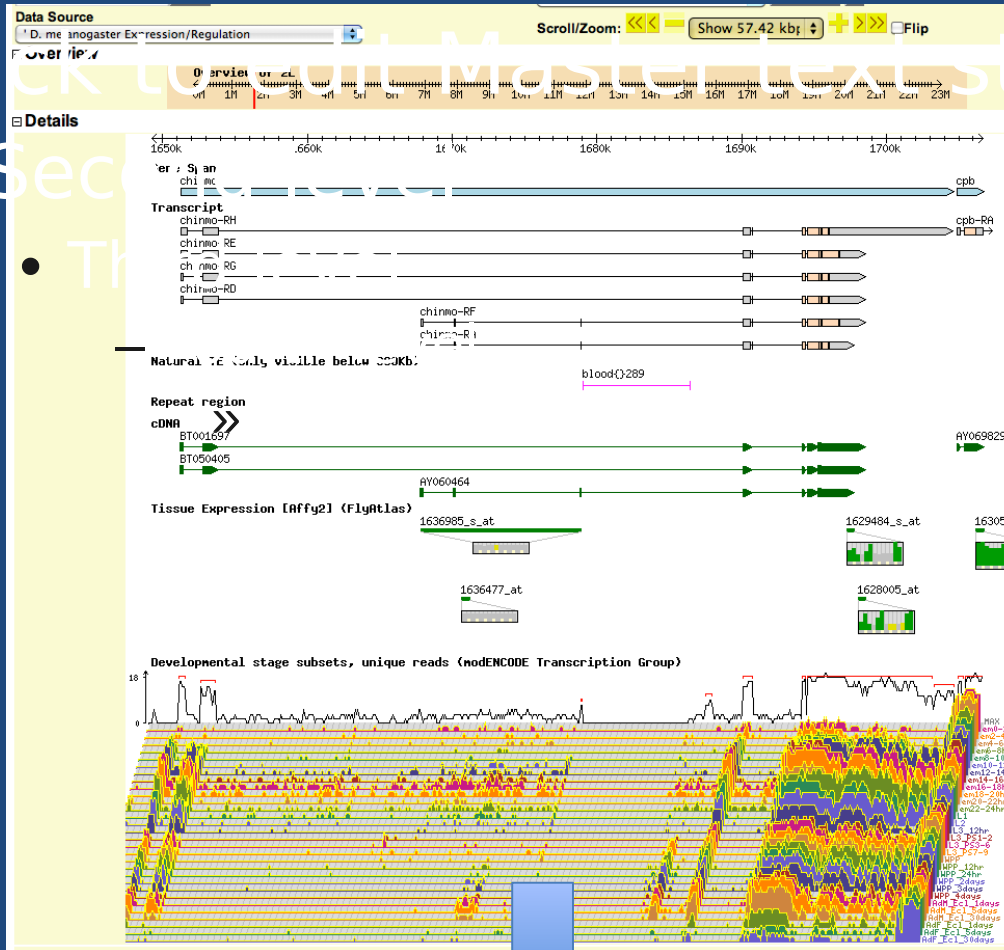
FlyAtlas tissue-level
expression data

modENCODE RNA-Seq
developmental
timecourse

GBrowse: Expression/Regulation

Click to edit Master text styles

- Second
- Third



gene
extents
annotated
transcripts
natural TE's

cDNAs

FlyAtlas tissue-level
expression data

modENCODE RNA-Seq
developmental
timecourse

GB: Expression/regulation data options

● Click to edit Master text styles

Clear highlighting Update Image

[\[Help and key for FlyBase Gene Report\]](#)

Tracks

- Reference Genome Annotations (Iso-1)** Select all Clear all
 - Gene Span CNS Repeat region
 - Transcript Nucleotide only (is below 100 bp)
- General** Select all Clear all
 - 3-frame translation (forward) DNA/GC Content 3-frame translation (reverse)
- Aligned Evidence** Select all Clear all
 - cDNA mRNA RNA-seq exon junctions
 - EST other aligned sequences PeptideAtlas peptides
- Previous Gene Models** Select all Clear all
 - Recently deleted transcripts
- Mapped Mutations** Select all Clear all
 - Transgenic Insertion Site
- Noncoding Features** Select all Clear all
 - Insulator class I Silencer TFBS - homeodomain Chromatin Domains (5-state model, Kc cells) Putative Brain Enhancers (Pfeiffer et al)
 - Insulator class II Regulatory Region TFBS - helix-loop-helix domain Chromatin Domains (9-state model, S2 cells) RNA Editing Sites
 - Protein Binding Site TFBS - HOT spot analysis TFBS - BTB/POZ domain Chromatin Domains (9-state model, BG3 cells)
 - Enhancer TFBS - zinc finger domain TFBS - other Origin of Replication
- Microarray Features** Select all Clear all
 - Affymetrix v1 Affymetrix v2
- Expression Levels: MicroArrays** Select all Clear all
 - Timecourse [cDNA] (Arbeitman et al.) Timecourse [amplicons] (Gauhar et al.) Tissue Expression [Affy2] (FlyAtlas)
- Expression Levels: RNA-seq** Select all Clear all
 - Developmental stage subsets (Baylor) Tissue culture cells (modENCODE Transcription Group) Treatments/Conditions, by strand
 - Developmental stage subsets, unique reads (modENCODE Transcription Group) Tissue culture cells, by strand (modENCODE Transcription Group)
- Expression Levels: RNA-Seq by Tissue** Select all Clear all
 - Digestive system Imaginal disc and other carcass Gonads and male accessory glands
 - Fat body and salivary glands CNS and adult head
- BLAST hit** Select all Clear all
 - Blast HSP

[Configure tracks...](#) Update Image

Select from Aligned Evidence: RNA-seq exon junctions

'Select all' from Expression Levels: RNA-Seq by Tissue

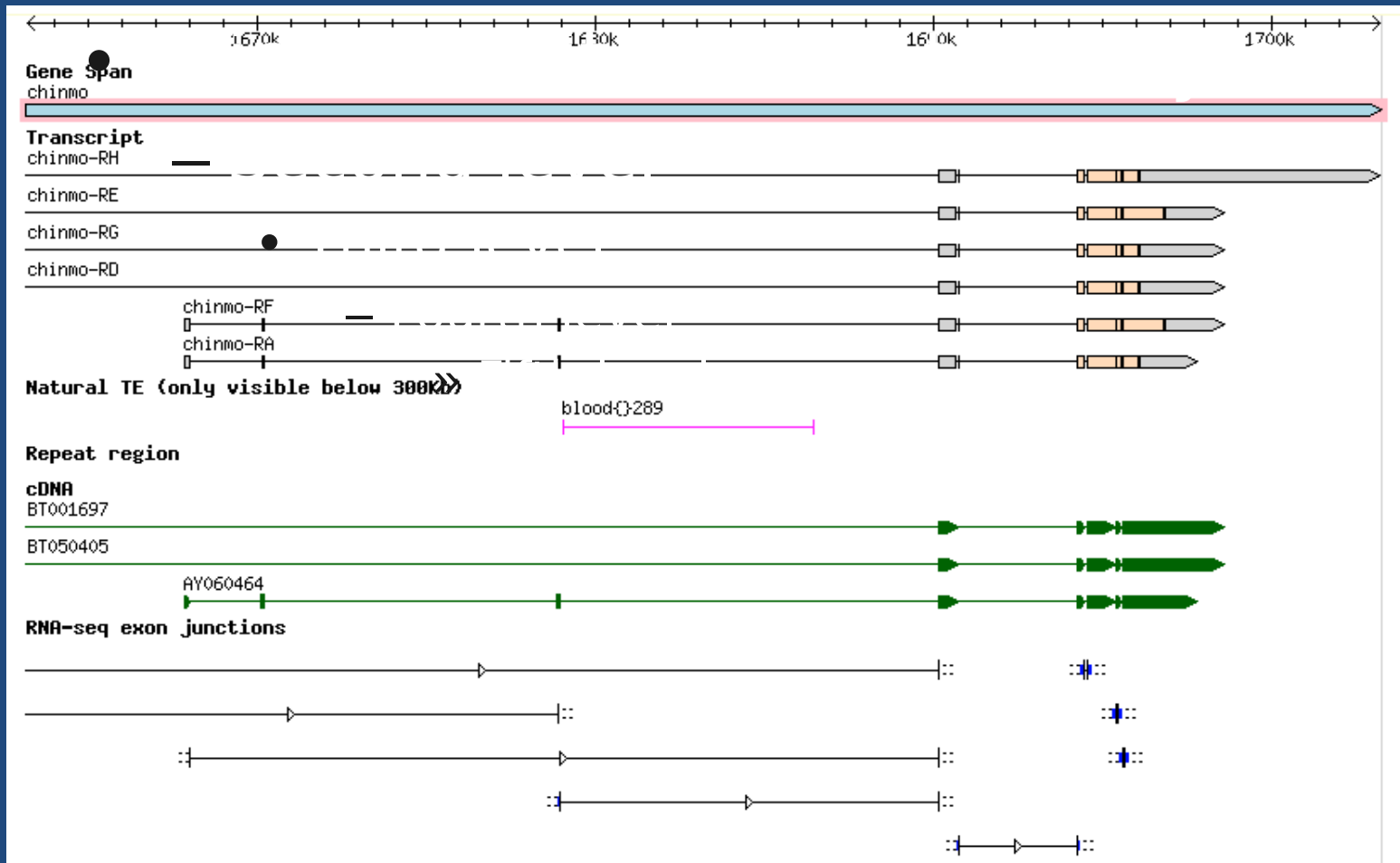
Click to edit Master text styles

Second level
Third level
Fourth level
Fifth level

<input type="checkbox"/> 3-frame translation (forward)			<input type="checkbox"/> DNA/GC Content	<input type="checkbox"/> 3-frame translation (reverse)
<input checked="" type="checkbox"/> cDNA			<input type="checkbox"/> mRNA	<input checked="" type="checkbox"/> RNA-seq exon junctions
<input type="checkbox"/> EST			<input type="checkbox"/> other aligned sequences	<input type="checkbox"/> PeptideAtlas peptides
<input type="checkbox"/> Timecourse [cDNA] (Arbeitman et al.)			<input type="checkbox"/> Timecourse [amplicons] (Gauhar et al.)	<input type="checkbox"/> Tissue Expression [Affy2] (FlyAtlas)
<input type="checkbox"/> Developmental stage subsets (Baylor)	<input type="checkbox"/> Tissue culture cells (modENCODE Transcription Group)	<input type="checkbox"/> Treatments/Conditions, by strand		
<input checked="" type="checkbox"/> Developmental stage subsets, unique reads (modENCODE Transcription Group)	<input type="checkbox"/> Tissue culture cells,by strand (modENCODE Transcription Group)			
<input checked="" type="checkbox"/> Digestive system	<input checked="" type="checkbox"/> Imaginal disc and other carcass	<input checked="" type="checkbox"/> Gonads and male accessory glands		
<input checked="" type="checkbox"/> Fat body and salivary glands	<input checked="" type="checkbox"/> CNS and adult head			

Configure tracks... Update Image

GBrowse: RNA-seq exon junctions



Sequence feature reports for RNA-Seq exon junctions

- **Click FlyBase** Master text styles

FB2012_01, released January 20th, 2012

Home Tools Files Species Documents Resources News Help Archives [Jump to Gene](#) [Go](#)

Profile Manager [Help](#) [Open All](#) [Close All](#)

General Information			
Symbol	Dmel:5:2L:1653835-1690117:+	Species	<i>D. melanogaster</i>
Feature type	exon_junction	FlyBase ID	FBsf0000101966
Collection	modENCODE_mRNA-Seq_U_junctions JCM_1_RNAseq_junctions	Associated gene(s)	chinmo

Genomic Location			
Chromosome (arm)	2L	Sequence location	2L:1,653,835..1,690,117 [+]

Map ([GBrowse](#)) [»»](#)

Recent Updates

Sequence Data

Length	
Comments	GT..AG
Sequence	

Experimental Data		
RNA-Seq Read Counts		
Library: BCM_1_RNAseq_junctions		
Library: modENCODE_mRNA-Seq_U_junctions		
modENCODE_mRNA-Seq_U_junctions	Expression stage(s)	Read Count
	embryonic stage 1 – 4	2
	embryonic stage 4 – 9	0
	embryonic stage 9 – 11	1
	embryonic stage 11 – 12	5
	embryonic stage 12 – 13	11
	embryonic stage 13 – 15	12
	embryonic stage 15 – 16	117
	embryonic stage 16	32
	embryonic stage 17(i)	147
	embryonic stage 17(ii)	195
	embryonic stage 17(iii)	48
	embryonic stage 17(iv)	181
	first instar larval stage	45
	second instar larval stage	52

Experimental Data

Collection Information

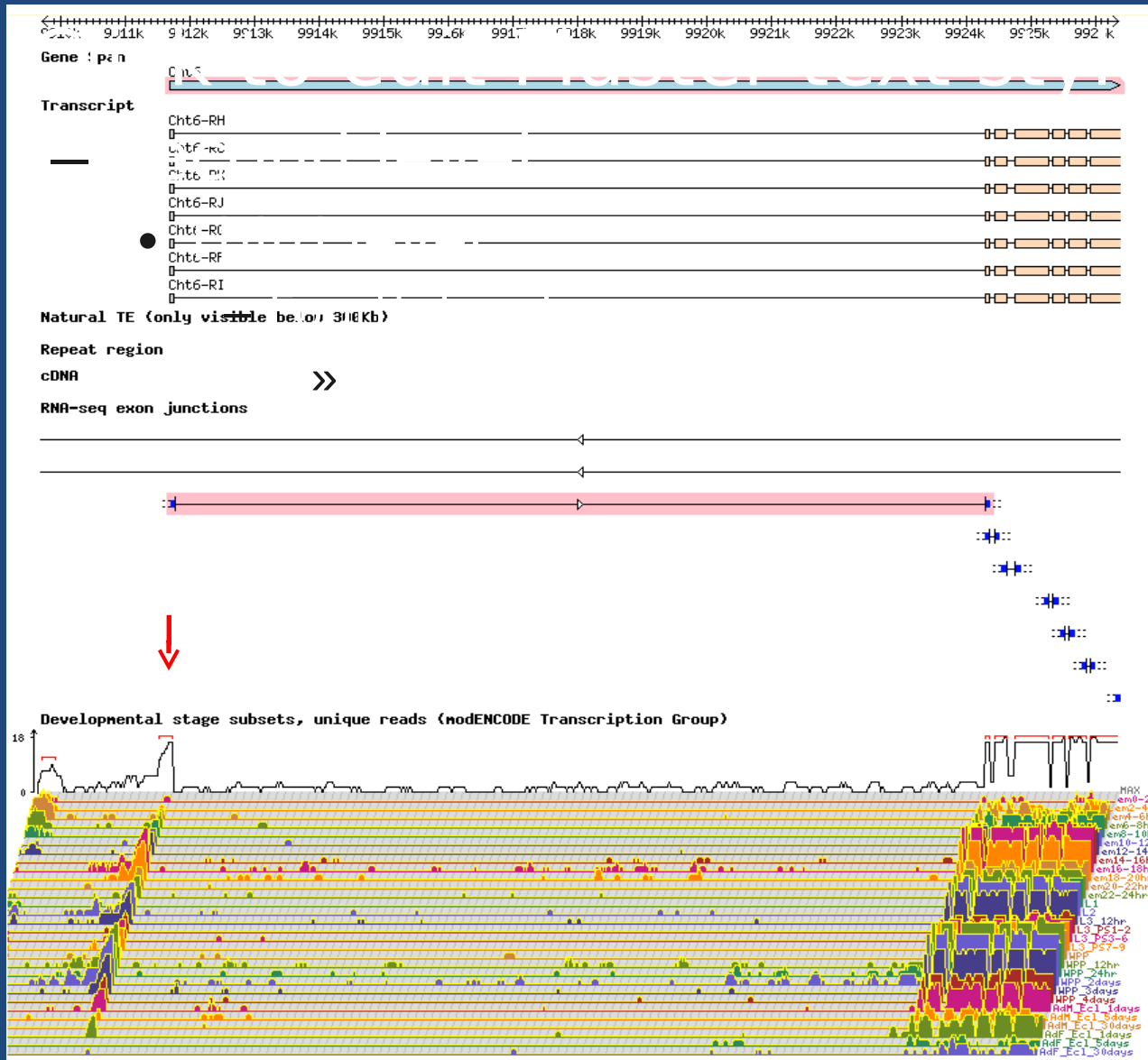
Collection Information

Collection: modENCODE_mRNA-Seq_U_junctions

Symbol

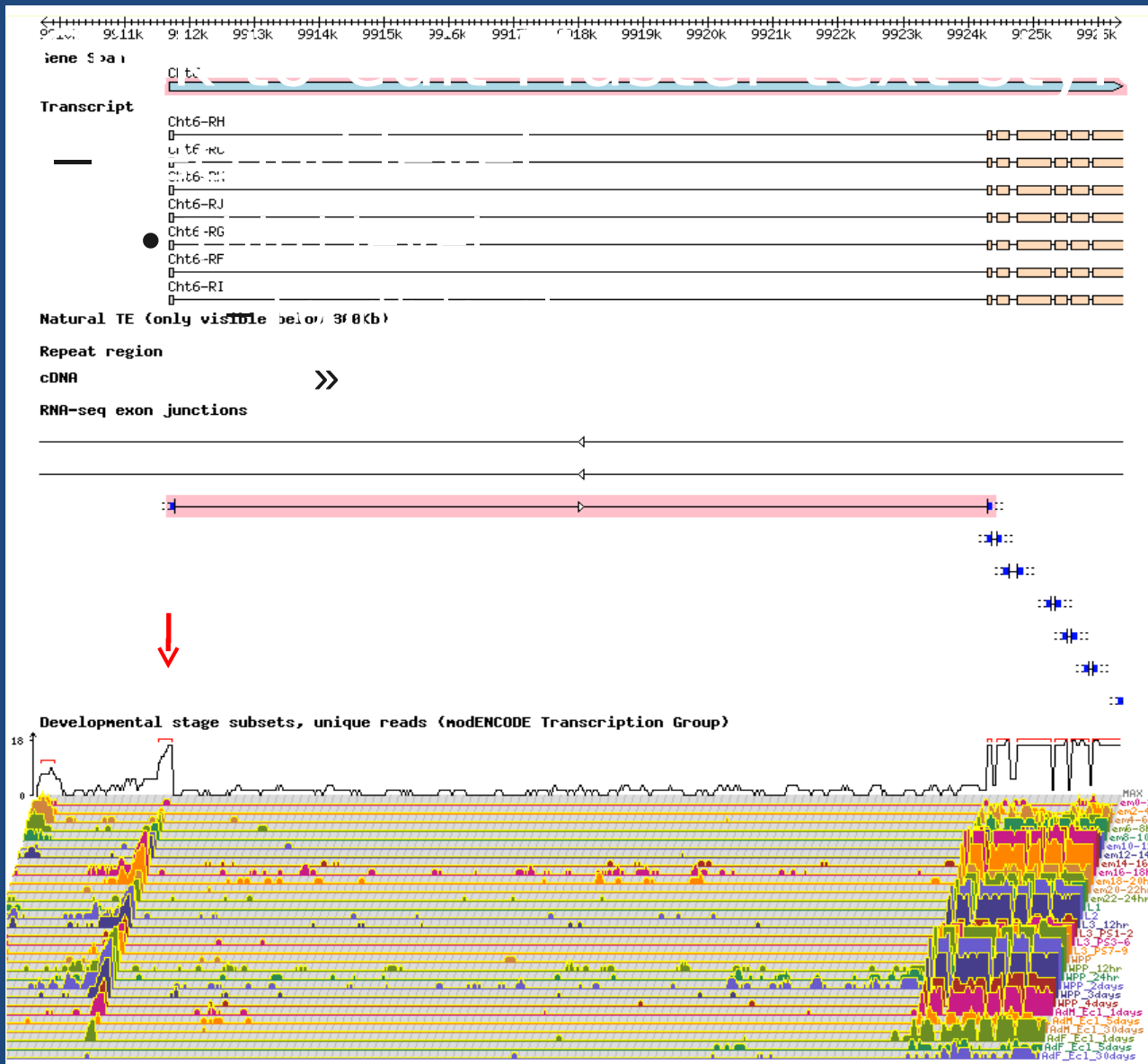
modENCODE_mRNA-Seq_U_junctions

New 5' exon (UTR) added to Cht6



S

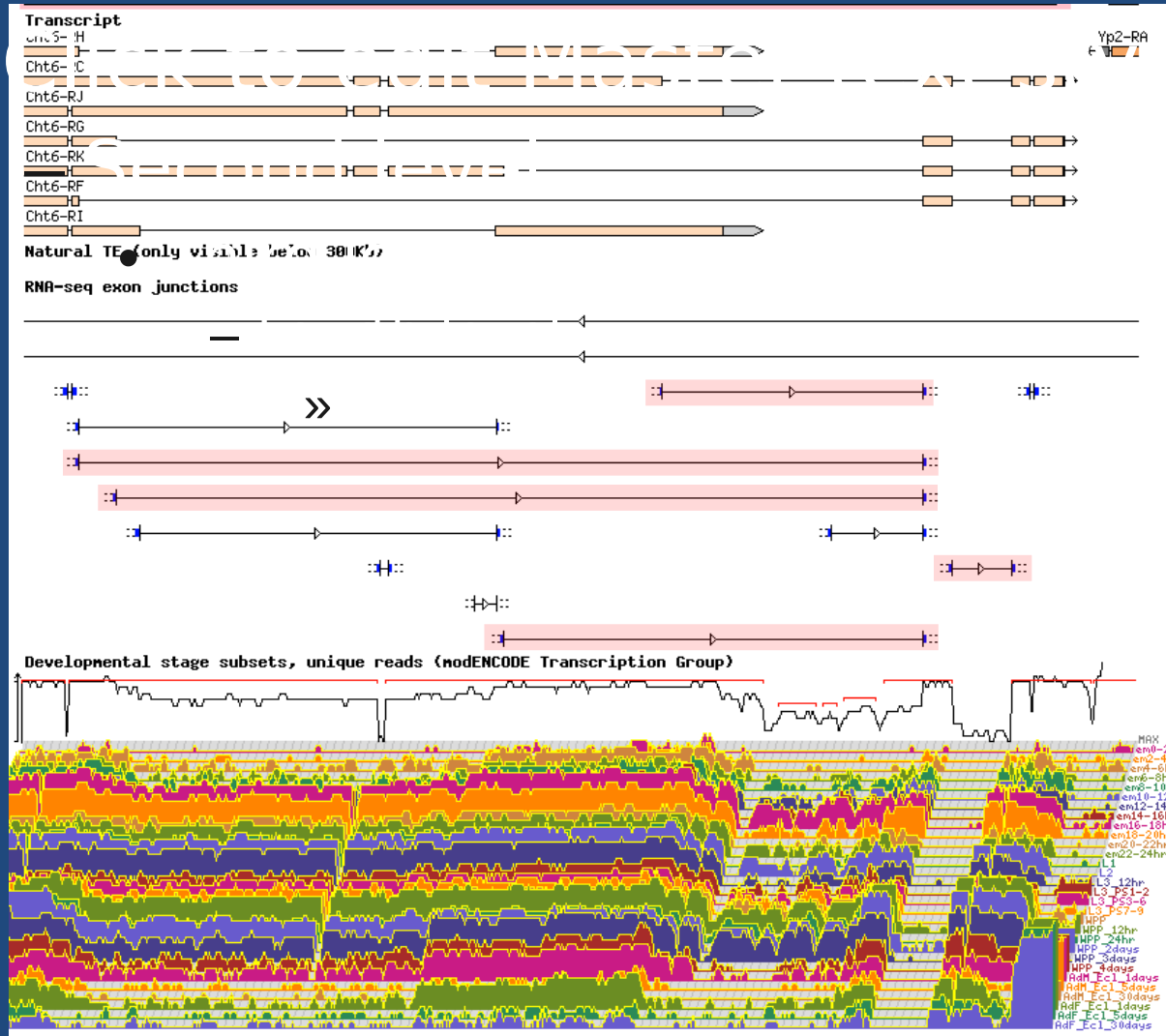
New 5' exon (UTR) added to Cht6



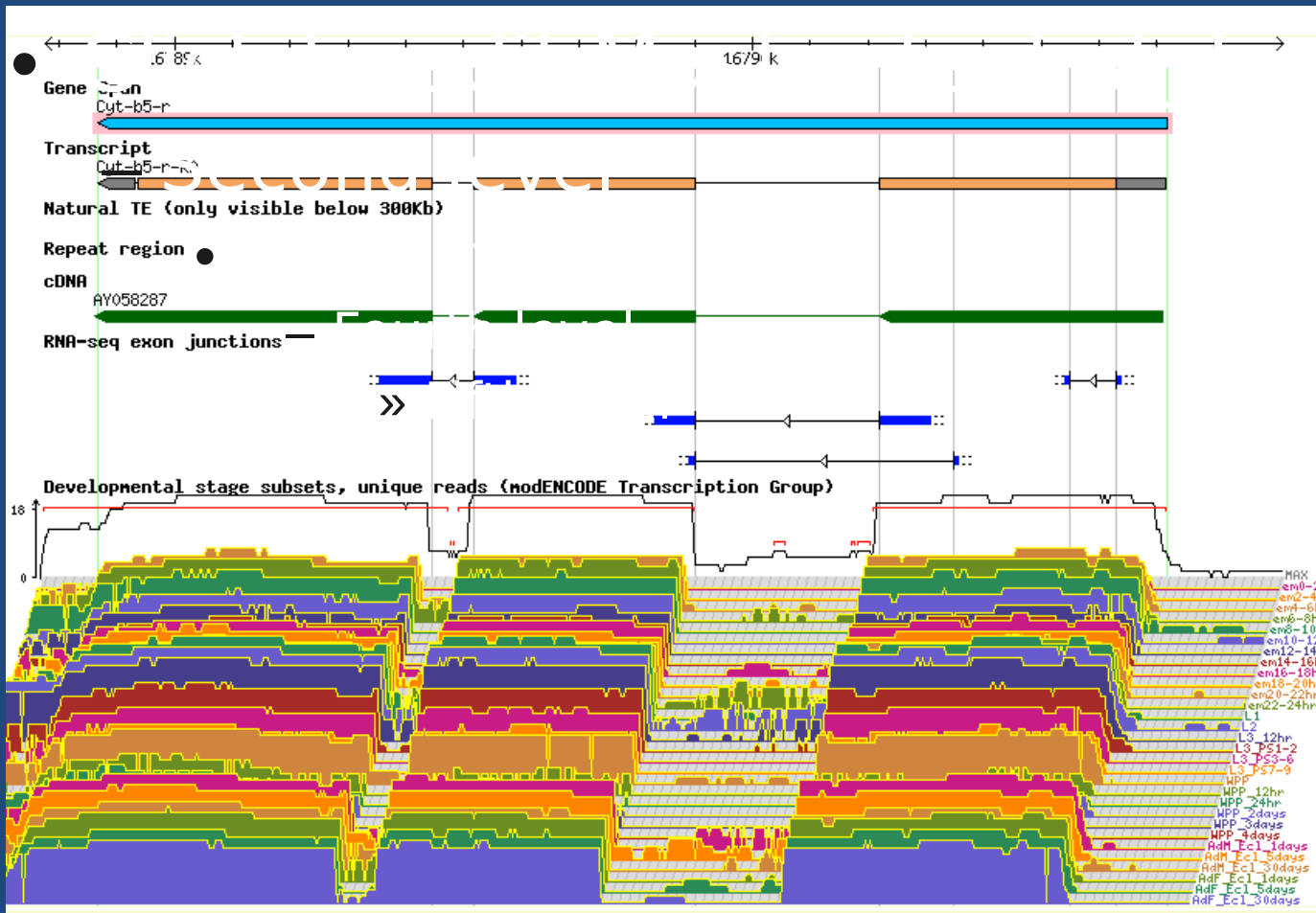
S



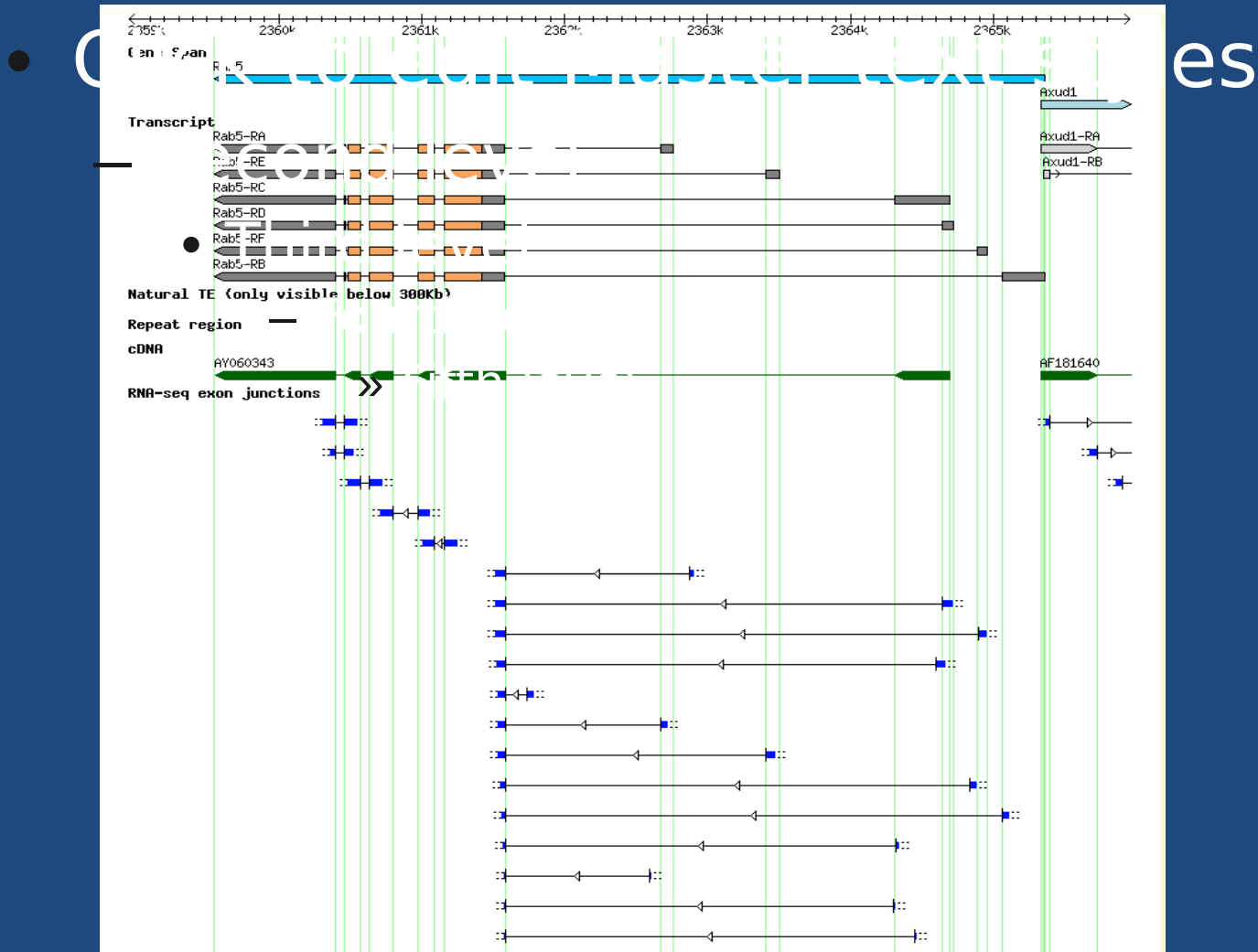
Gene merge: Cht6 + downstream genes



Junctions: low frequency junctions may not be annotated



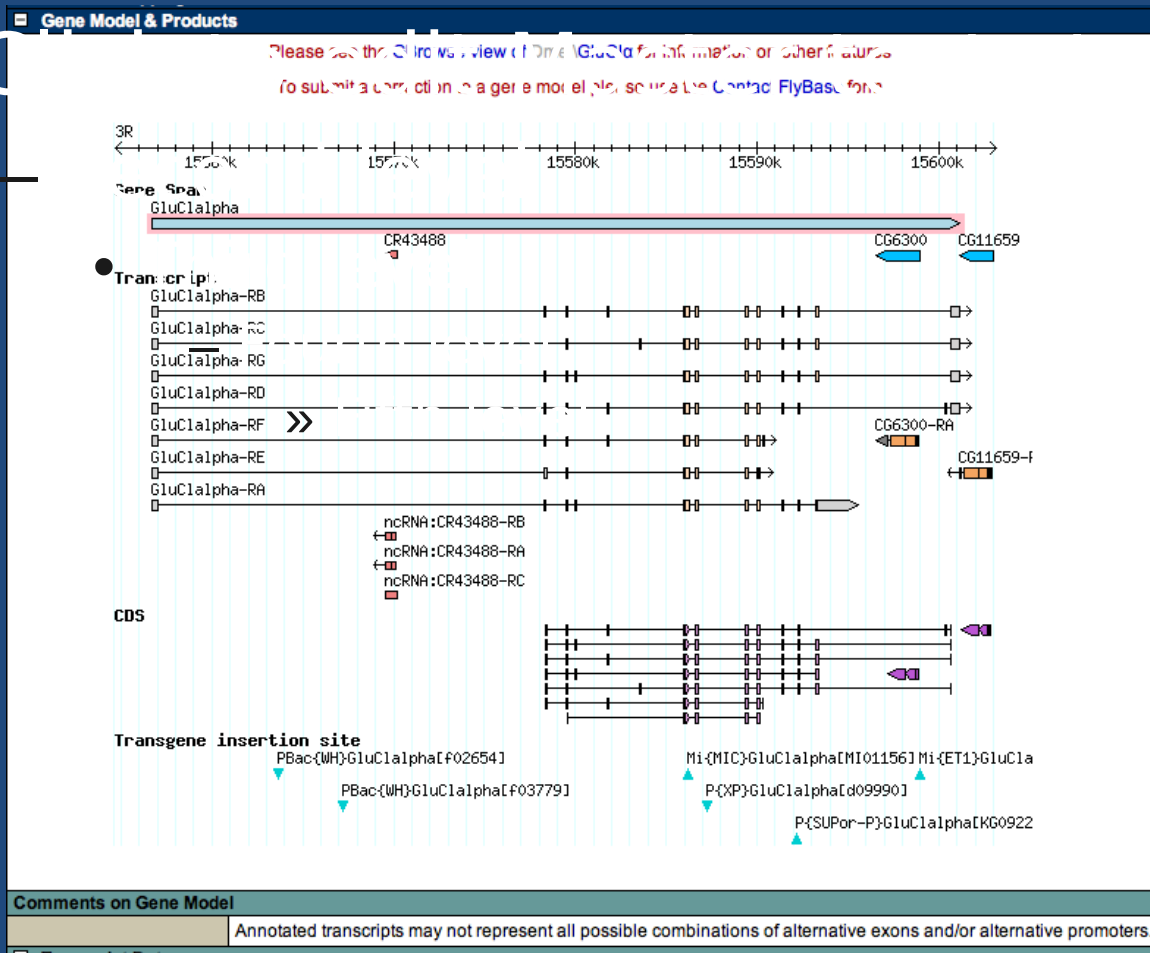
Junctions: variations in 5' UTR may not be annotated



Gene model comments

- Low-frequency RNA-Seq exon junction(s) not annotated.
- Annotated transcripts do not represent all supported alternative splices within 5' UTR.

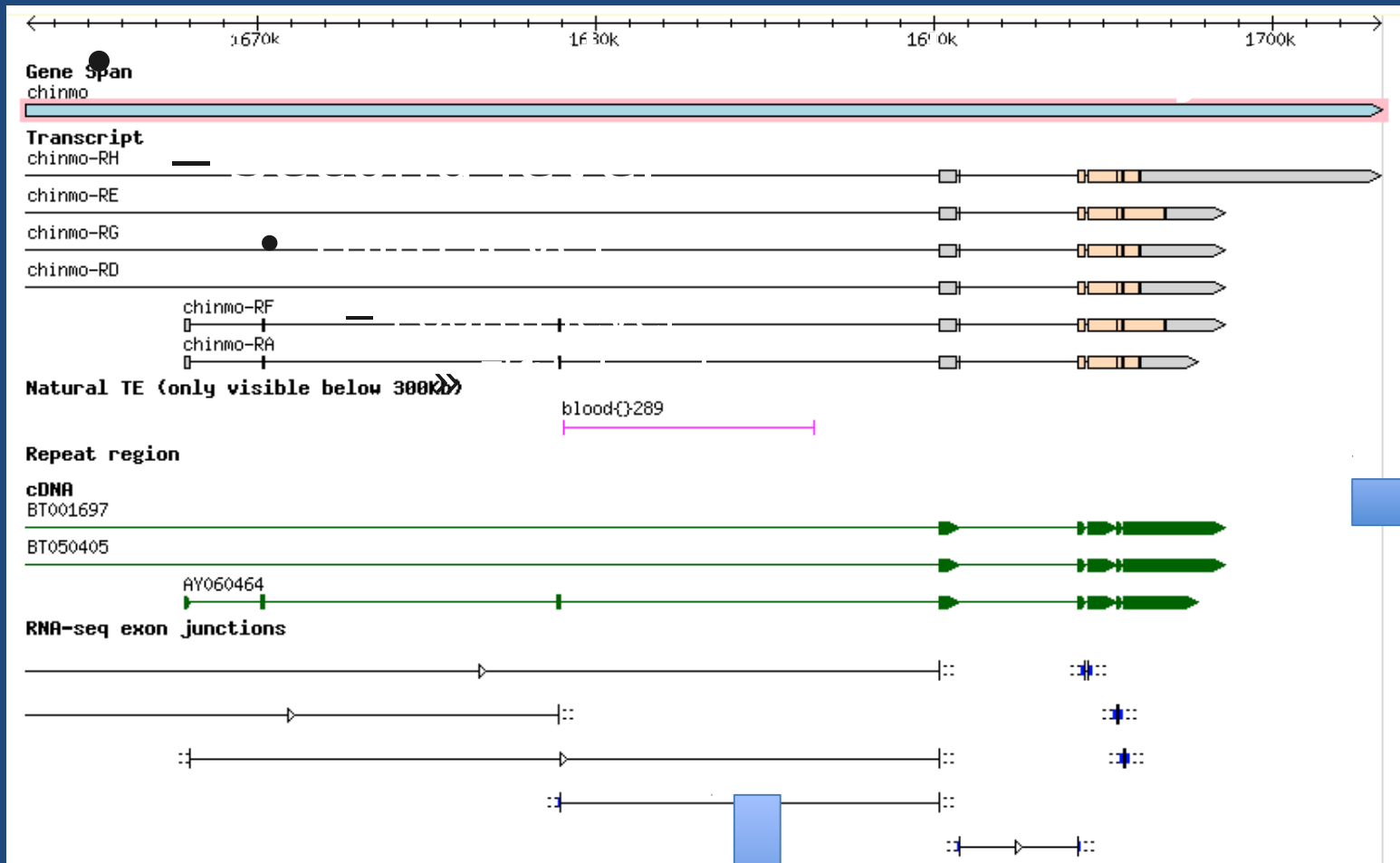
Comments on Gene Model (permutation problem)



es

Annotated transcripts may not represent all possible combinations of alternative exons and/or alternative promoters.

GBrowse: expression data

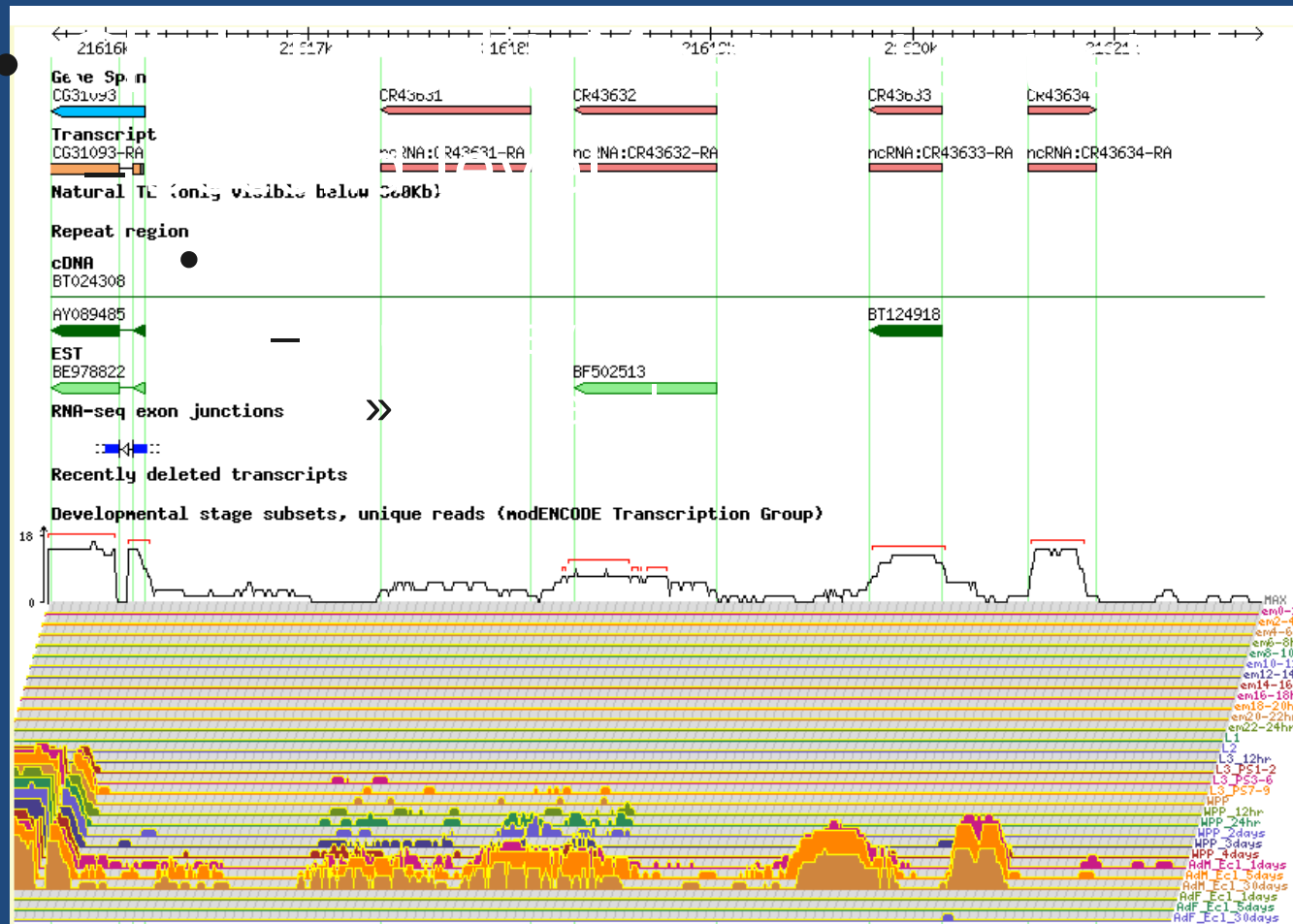




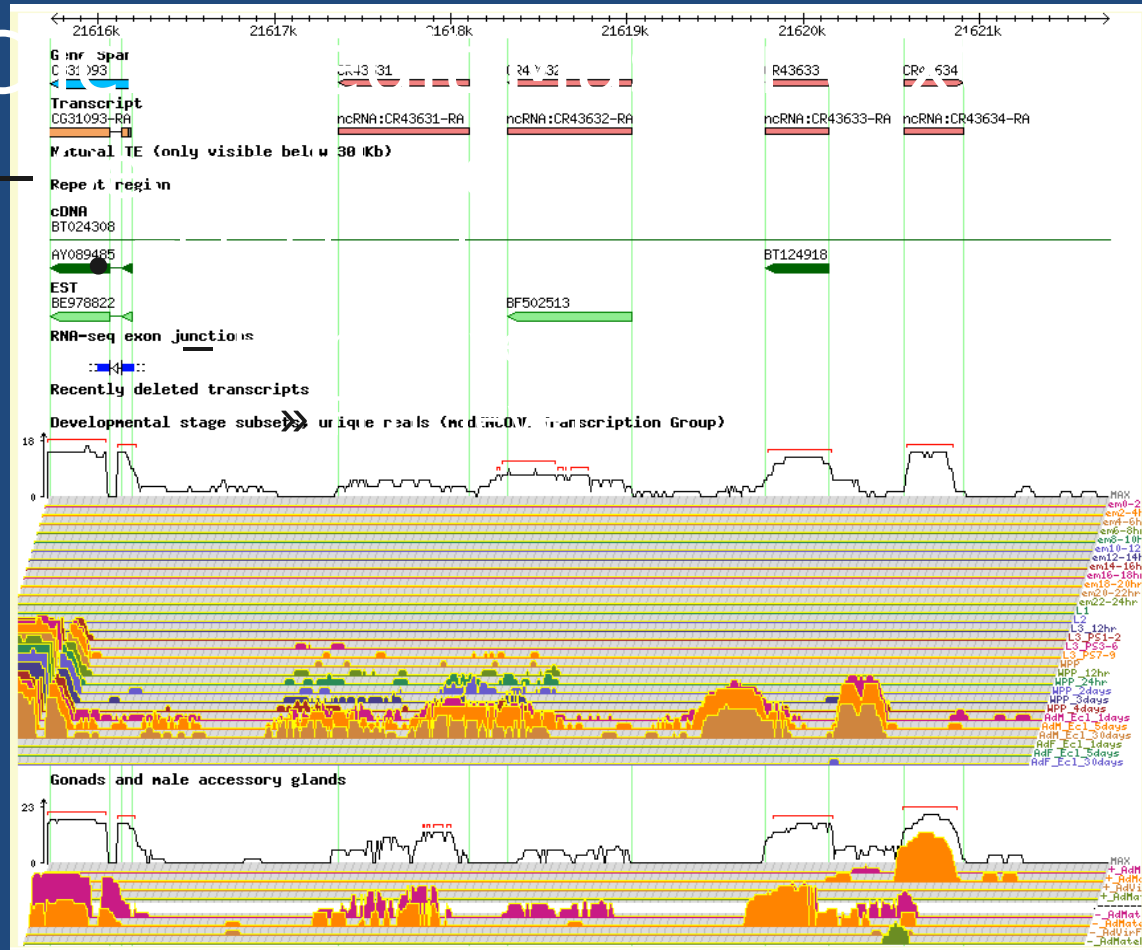
ES

strand-specific
RNA-Seq data
from isolated
L/A tissues

New genes: ncRNA genes

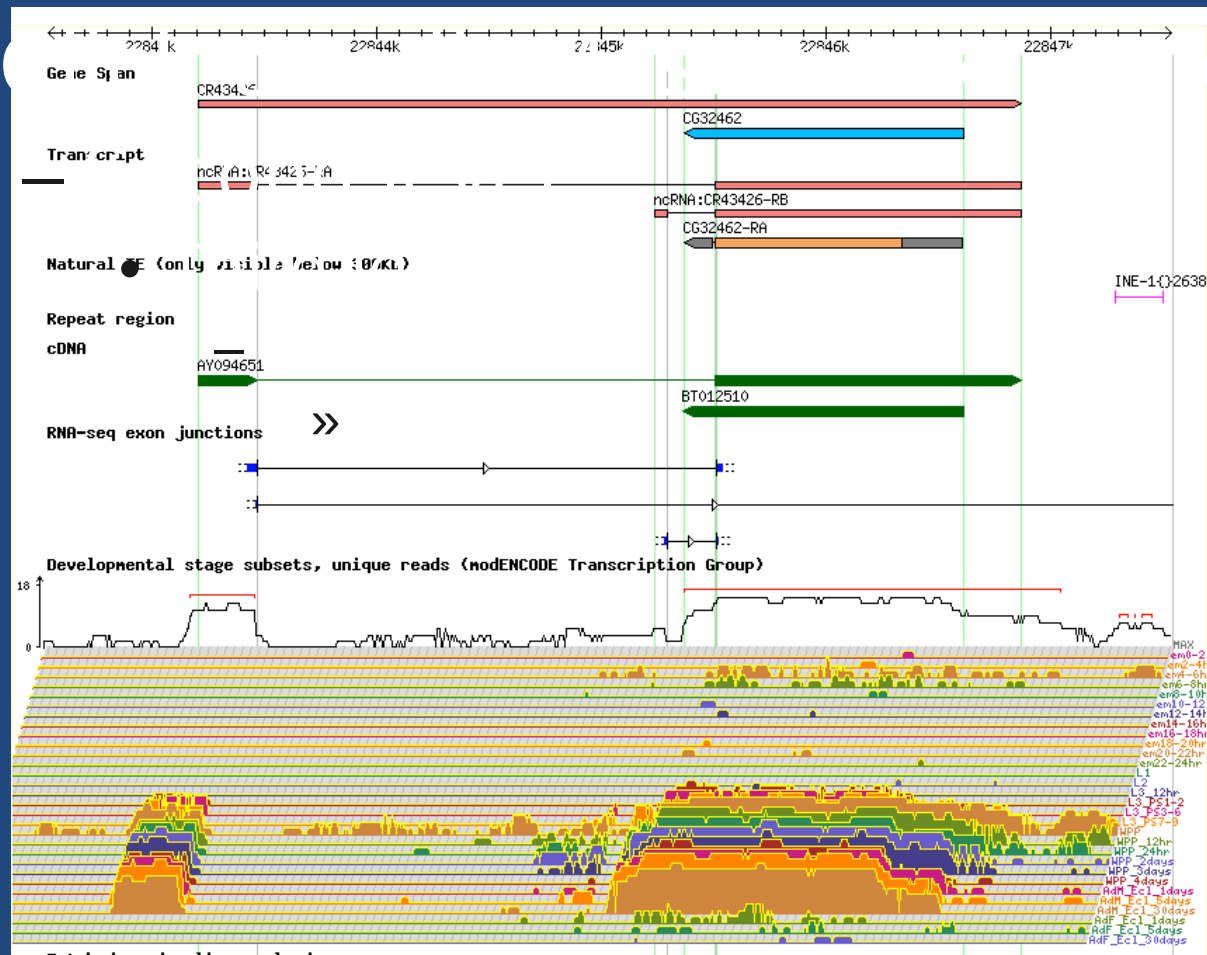


Stranded RNA-Seq data: ncRNA genes



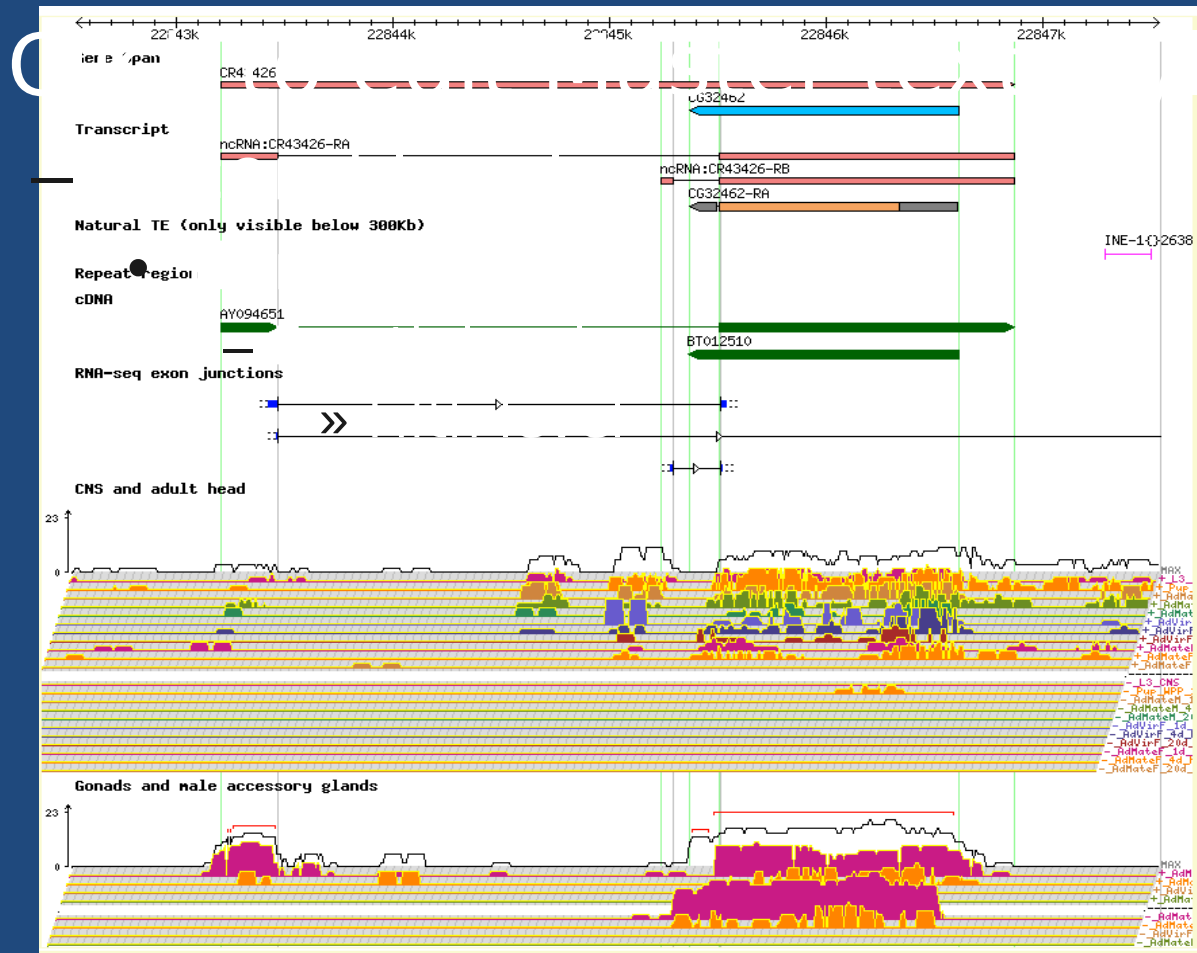
es

New genes: anti-sense RNAs



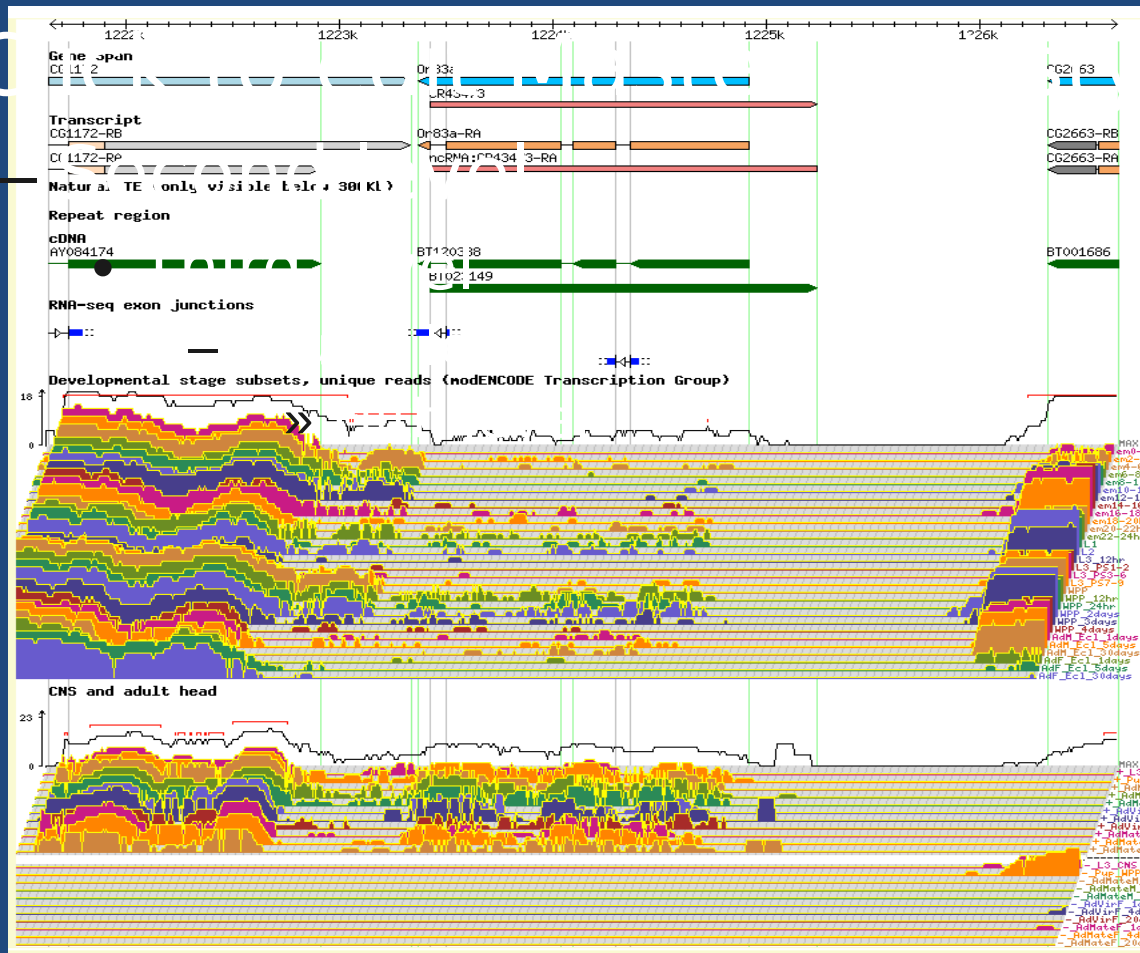
es

Stranded RNA-Seq data: anti-sense RNAs



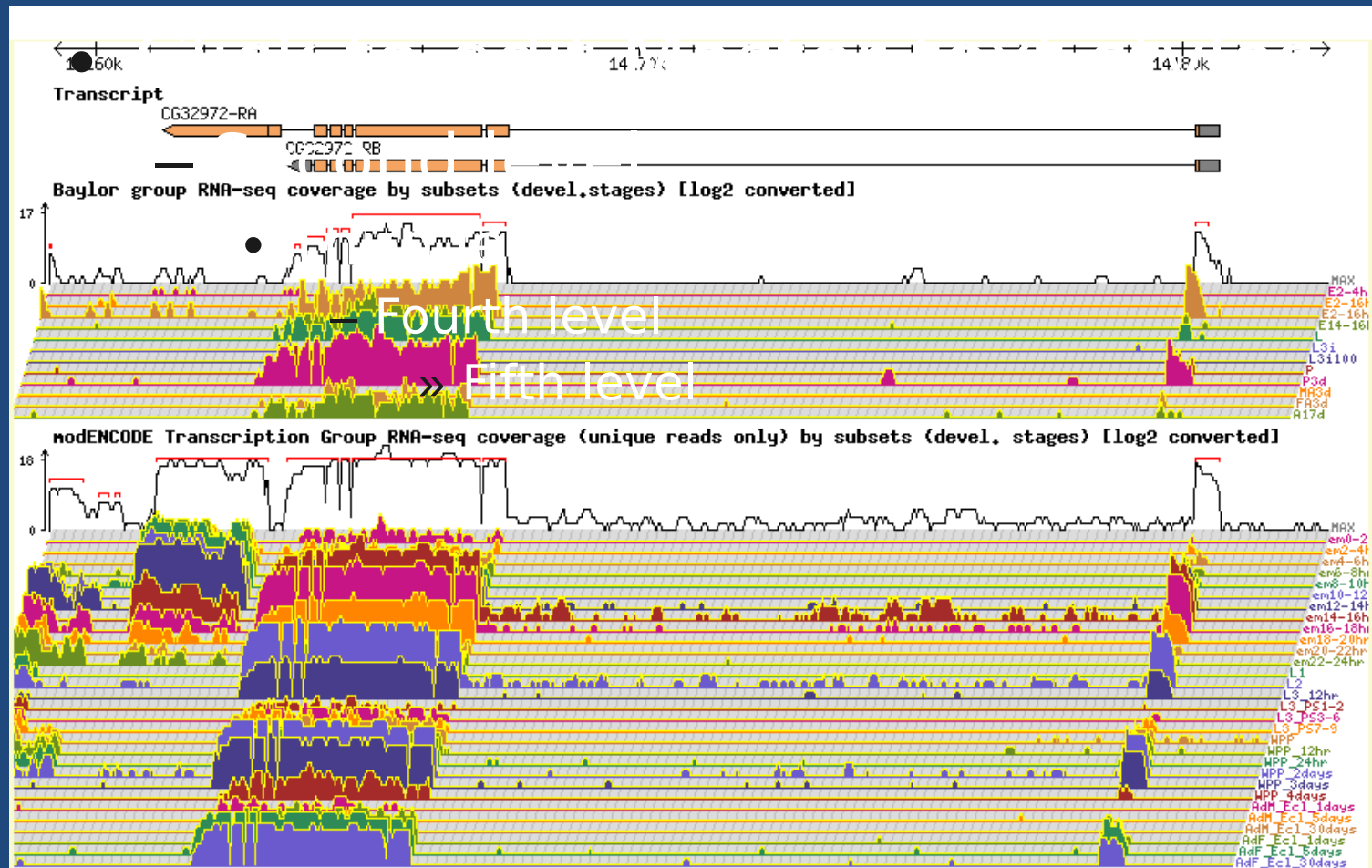
es

Stranded RNA-Seq data: anti-sense RNAs


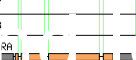


es

Gene split: CG32972



Intermediate 3' UTR extensions not annotated

- Click  styles
- Sec 
- T 



Gene model comments

- RNA-Seq data support multiple isoforms with extended 3' UTRs of differing length, some of which appear to be stage and/or tissue specific.

Impact of RNA-Seq

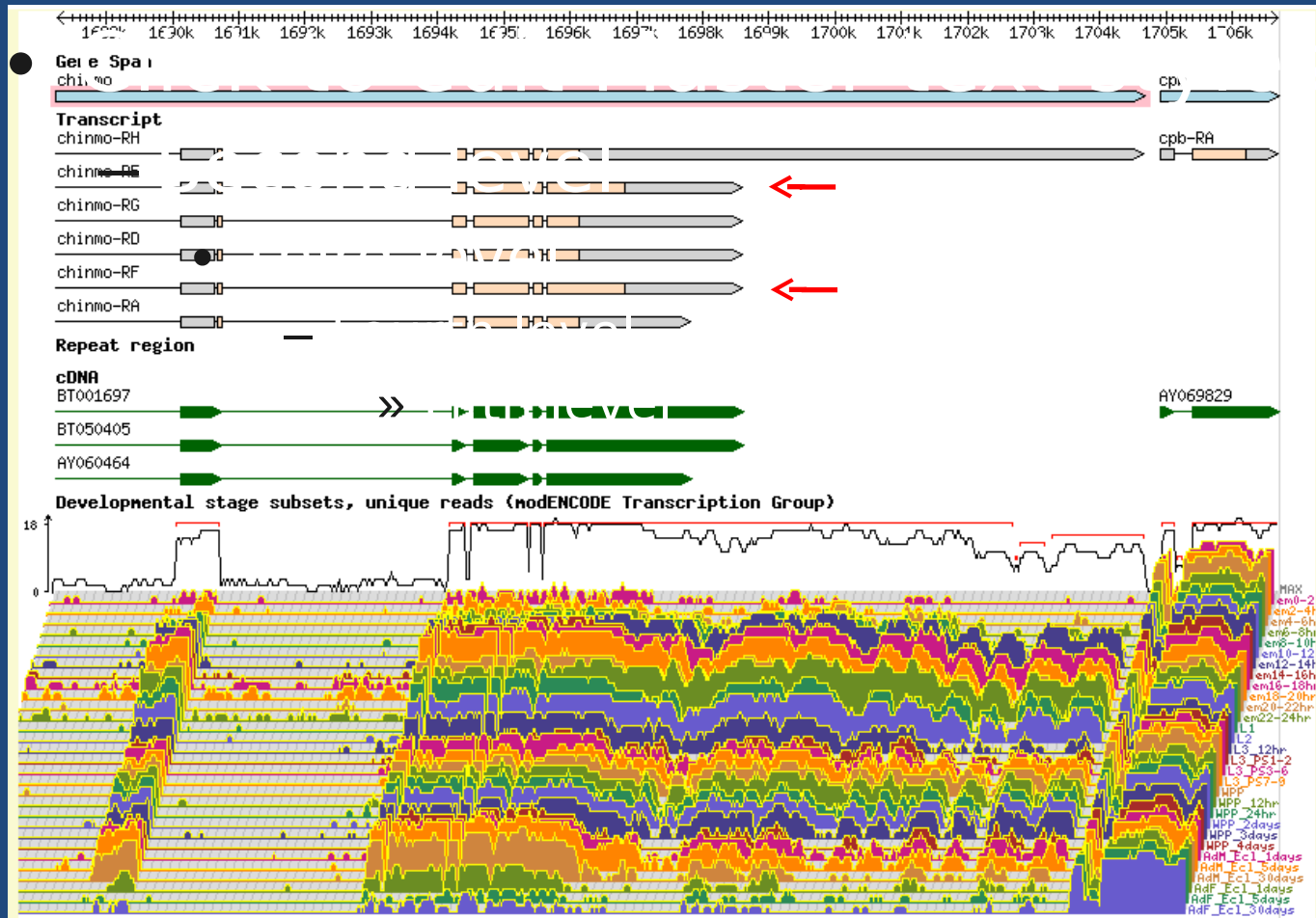
- ^{data} New and improved annotations:
 - New genes, coding and non-coding
 - New 5' exons/promoters
 - New internal exons
 - Extended 3' UTRs
 - Merges & splits

Impact of RNA-Seq

data

- Major source of reannotation flags:
 - Unannotated junctions (complete categorized and prioritized list)
 - Targeted MIP cDNAs
 - RNA-Seq coverage data without corresponding annotation

New data: stop-codon readthrough



Gene model comments

- gene_with_stop_codon_read_through
; SO:0000697
- Stop-codon suppression (UGA)
postulated; FBrf0216884.

Transcript features annotated on the genome: RNA editing sites

Genome browser interface showing transcript features and RNA editing sites. The top track displays the gene span and transcripts (eag-RB, eag-RN, eag-RG, eag-RD) with exons and introns. Below, RNA editing sites are marked with blue arrows and 'A' labels. The interface includes a 'Tracks' panel with various annotation categories:

- Reference Genome Annotations (Iso-1)
 - Gene Span
 - Transcript
 - CDS
 - Natural TE (only visible below 300Kb)
 - Repeat region
- General
- Aligned Evidence
- Previous Gene Models
- Mapped Mutations
 - Insulator class I
 - Insulator class II
 - Protein Binding Site
 - Enhancer
 - Silencer
 - Regulatory Region
 - TFBS - HOT spot analysis
 - TFBS - zinc finger domain
 - TFBS - homeodomain
 - TFBS - helix-loop-helix domain
 - TFBS - BTB/POZ domain
 - TFBS - other
 - Chromatin Domains (5-state model, Kc cells)
 - Chromatin Domains (9-state model, S2 cells)
 - Chromatin Domains (9-state model, BG3 cells)
 - Origin of Replication
 - Putative Brain Enhancers (Pfeiffer et al)



Sequence Feature: modified_RNA_base_feature

Home Tools Files Species Documents Resources News Help Archives [Jump to Gene](#) [Go](#)

Profile Manager [+](#) [-](#) [?](#) [Help](#) [Open All](#) [Close All](#)

General Information

Symbol	A-I_edit_000938	Species	<i>D. melanogaster</i>
Feature type	modified_RNA_base_feature	FlyBase ID	FB:mf000938:02
Collection	mE_A-to-I_RNA_Editing_Sites	Associated gene(s)	

Genomic Location

Chromosome (arm)	X	Sequence location	X:14,898,809..14,898,809 [+]
------------------	---	-------------------	------------------------------

Sequence Data

Length	
Comments	Site of A-to-I editing in CDS. A to Asp substitution.
Sequence	

Associated Information

Gene(s) (targeted or local)	eag
Allele(s)	»
Transcripts(s)	eag-RA eag-RB eag-RD eag-RG

Experimental Data

RNA Editing Sites

Library: mE_A-to-I_RNA_Editing_Sites

mE_A-to-I_RNA_Editing_Sites	Expression stage(s)	Read Count	Editing Frequency
	embryonic stage 1 – 4	0	N/A
	embryonic stage 4 – 9	0	N/A
	embryonic stage 9 – 11	3	0.00
	embryonic stage 11 – 12	3	0.00
	embryonic stage 12 – 13	0	N/A
	embryonic stage 13 – 15	0	N/A
	embryonic stage 15 – 16	3	0.00
	embryonic stage 16	9	0.00
	embryonic stage 17(i)	12	0.00
	embryonic stage 17(ii)	42	0.12
	embryonic stage 17(iii)	8	0.00
	embryonic stage 17(iv)	22	0.00

Click to master
Second level
Fourth level

modENCODE TSS: FlyBase GBrowse

Instructions [\[Help\]](#) [\[Reset\]](#) [\[Bookmark this\]](#)

Search using a sequence name or range, gene symbol (but not full name, for example, ct but not cut), gene identifier, insertion symbol, insertion identifier, or other landmark. Searches are case-sensitive. Wildcards are not supported at this time. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: [cnn](#), [FBgn000490](#), [X:200000..220000](#), [2L:80,000..100,000](#), [2R:80,000..100,000](#), [3L:80,000..100,000](#), [3R:80,000..100,000](#), [4:100000..120000](#).

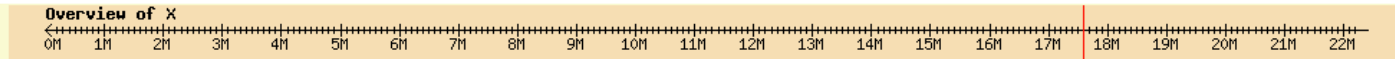
Search

Landmark or Region:

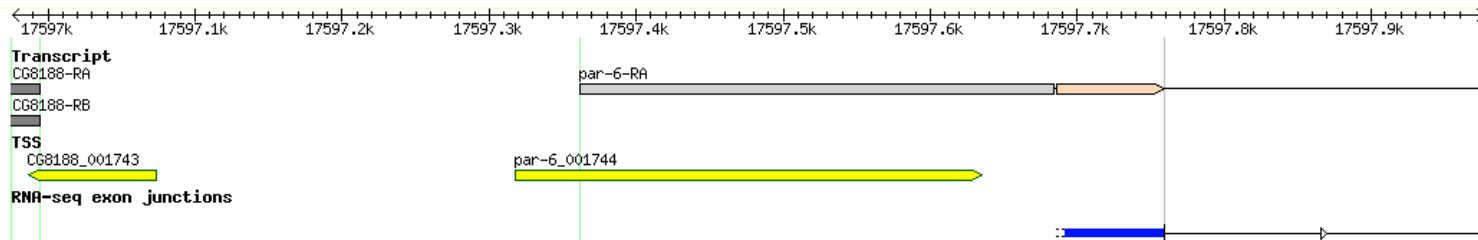
X:17596976..17597975

Data Source

Overview



Details



[Clear highlighting](#)

[\[Help and key for Stock Center Data evidence tiers\]](#)

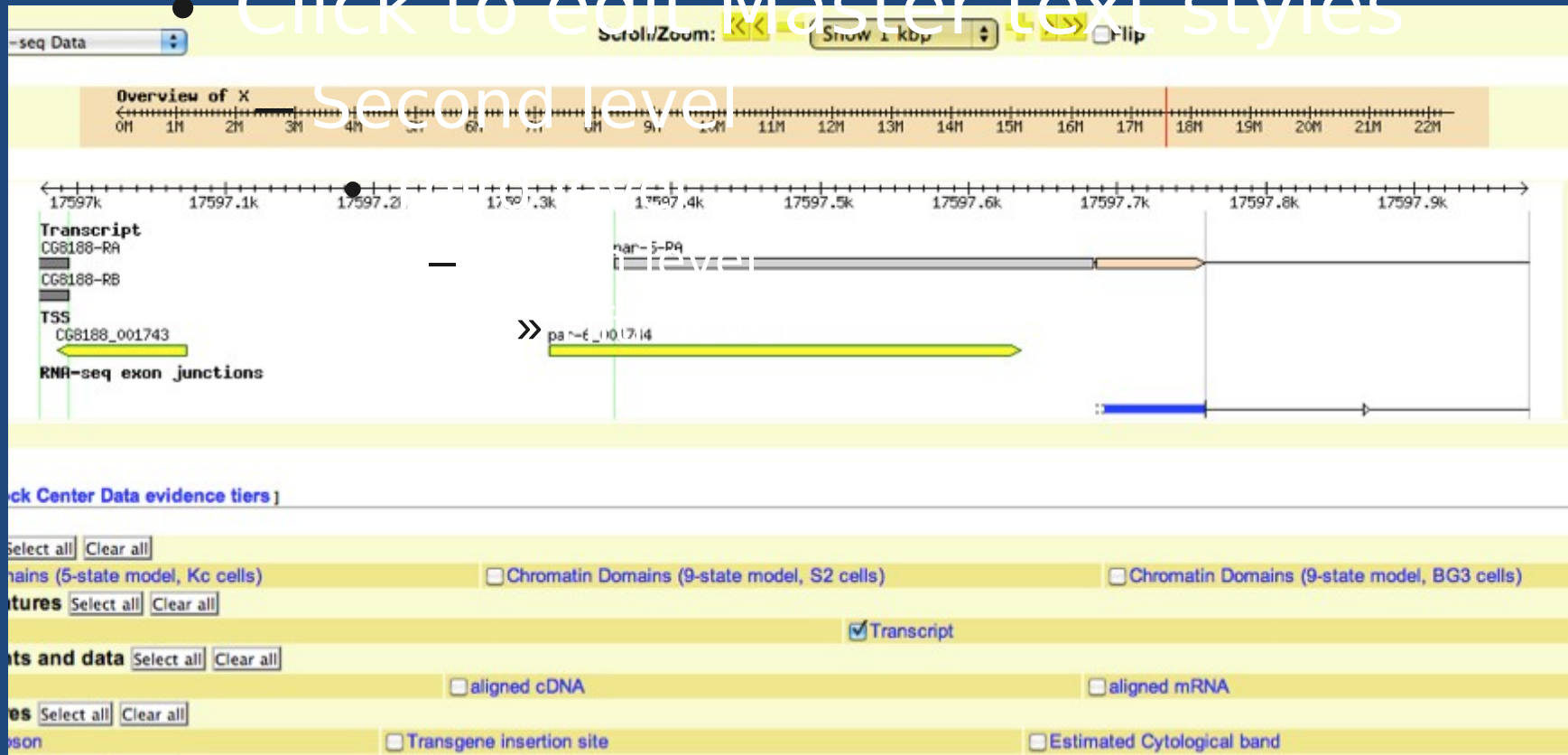
Tracks

- Chip-seq data**
 - Chromatin Domains (5-state model, Kc cells)
 - Chromatin Domains (9-state model, S2 cells)
 - Chromatin Domains (9-state model, BG3 cells)
- Gene Model features**
 - Gene Span
 - Transcript
- Genome reagents and data**
 - aligned EST
 - aligned cDNA
 - aligned mRNA
- Genomic features**
 - Natural transposon
 - Transgene insertion site
 - Estimated Cytological band
- Mapped features**
 - Insulator class I
 - Insulator class II
- Other RNA-seq data**
 - cell lines expression (by strand) [log2 converted]
 - expression under treatments (by strand) [log2 converted]
- RNA-seq data**
 - RNA-seq exon junctions
 - modENCODE Transcription Group RNA-seq coverage (unique reads only) by subsets (devel. stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage (embryonic stages, by strand) [log2 converted]
 - K. White group RNA-seq coverage by subsets (devel.stages) [log2 converted]
 - Baylor group RNA-seq coverage by subsets (devel.stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage (all reads) by subsets (devel. stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage by subsets (tissue culture cells) [log2 converted]
- RNA-seq data, expression by tissue**
 - digestive system (by strand) [log2 converted]
 - imaginal disc and other carcass (by strand) [log2 converted]
 - gonads and male accessory glands (by strand) [log2 converted]
 - fat body and salivary glands (by strand) [log2 converted]
 - CNS and adult head (by strand) [log2 converted]
- External Annotation Tracks**

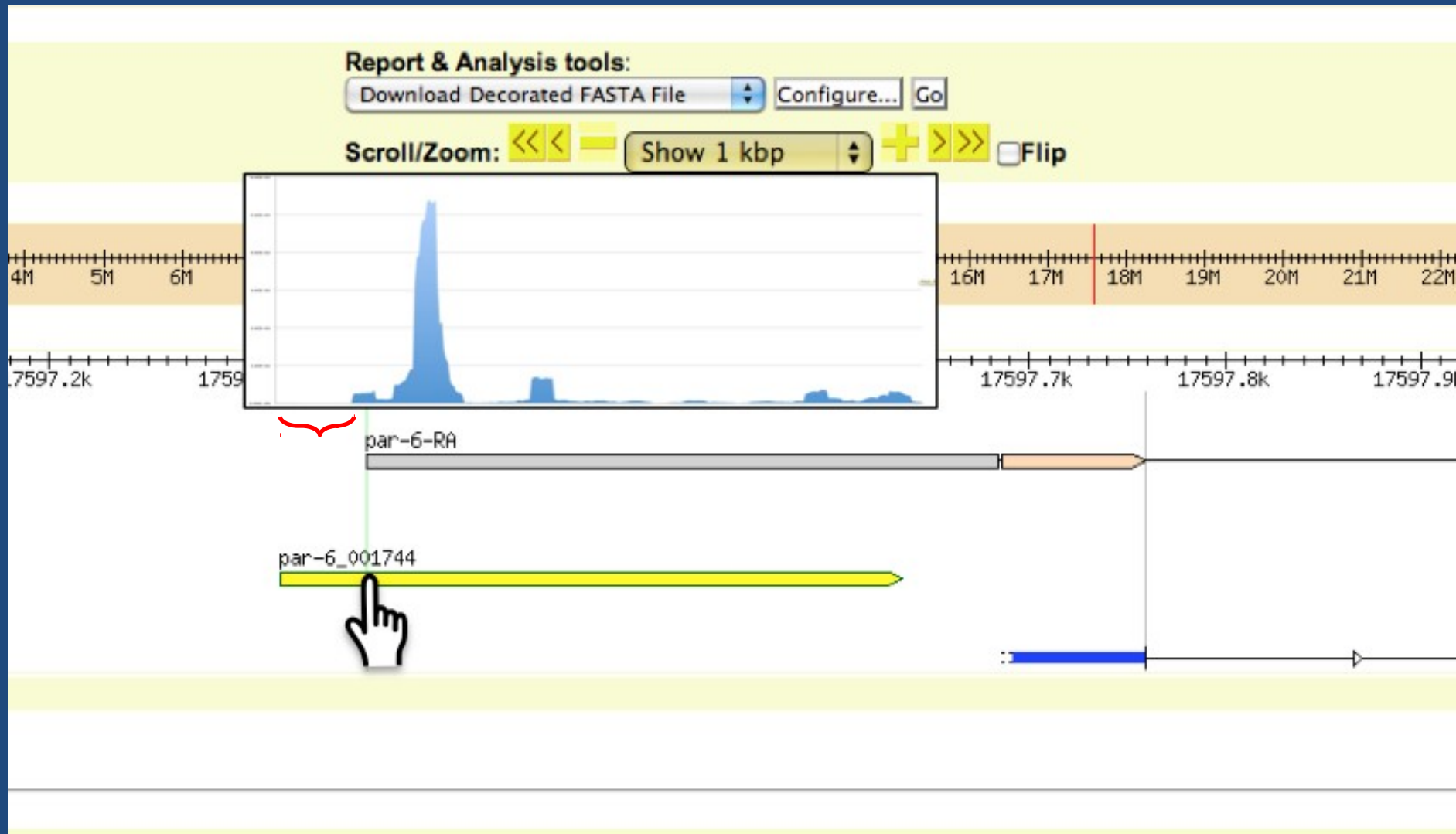
Transcription start sites
range represented as extent on genome

Click to edit Master text styles

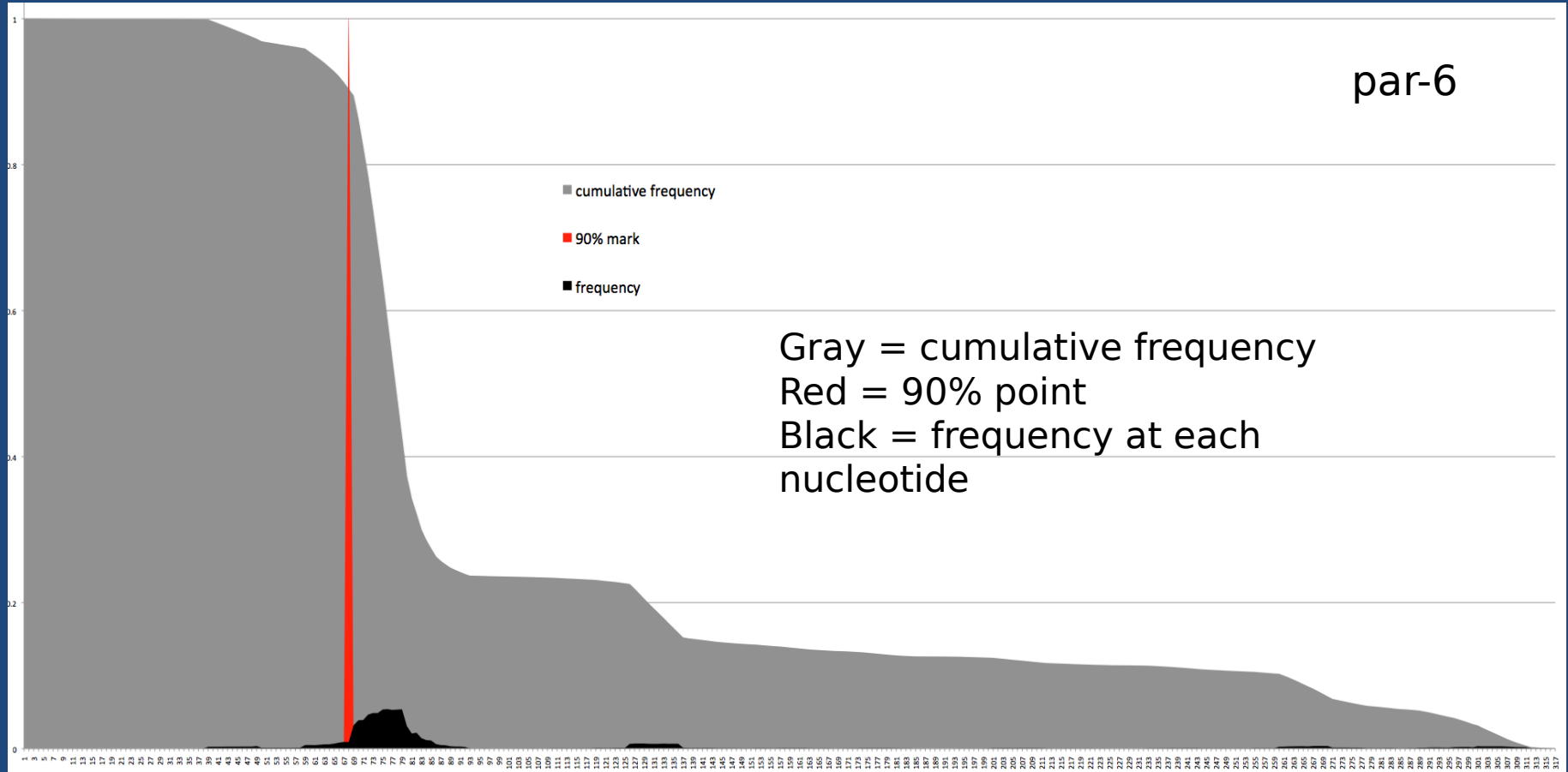
Second level



TSS data: frequency distributions

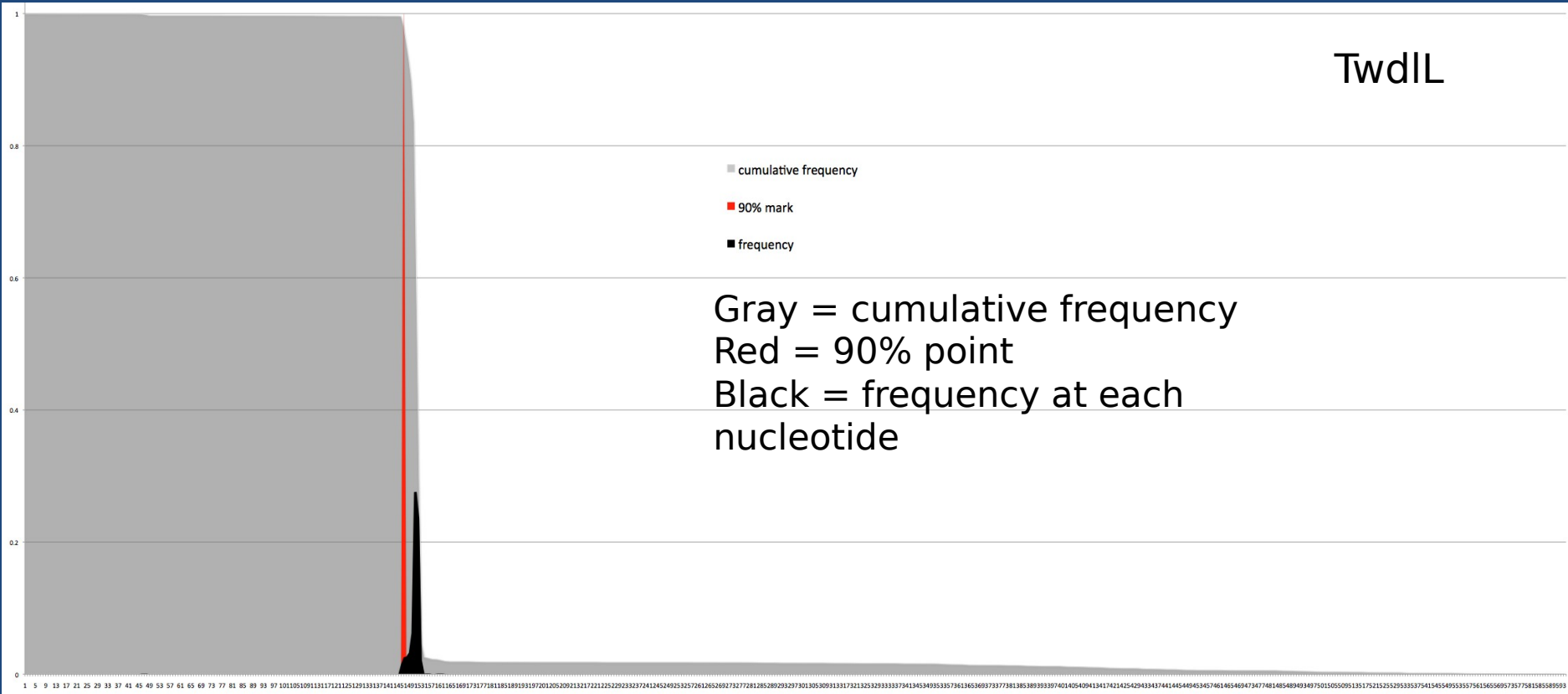


TSS annotation at 90% point

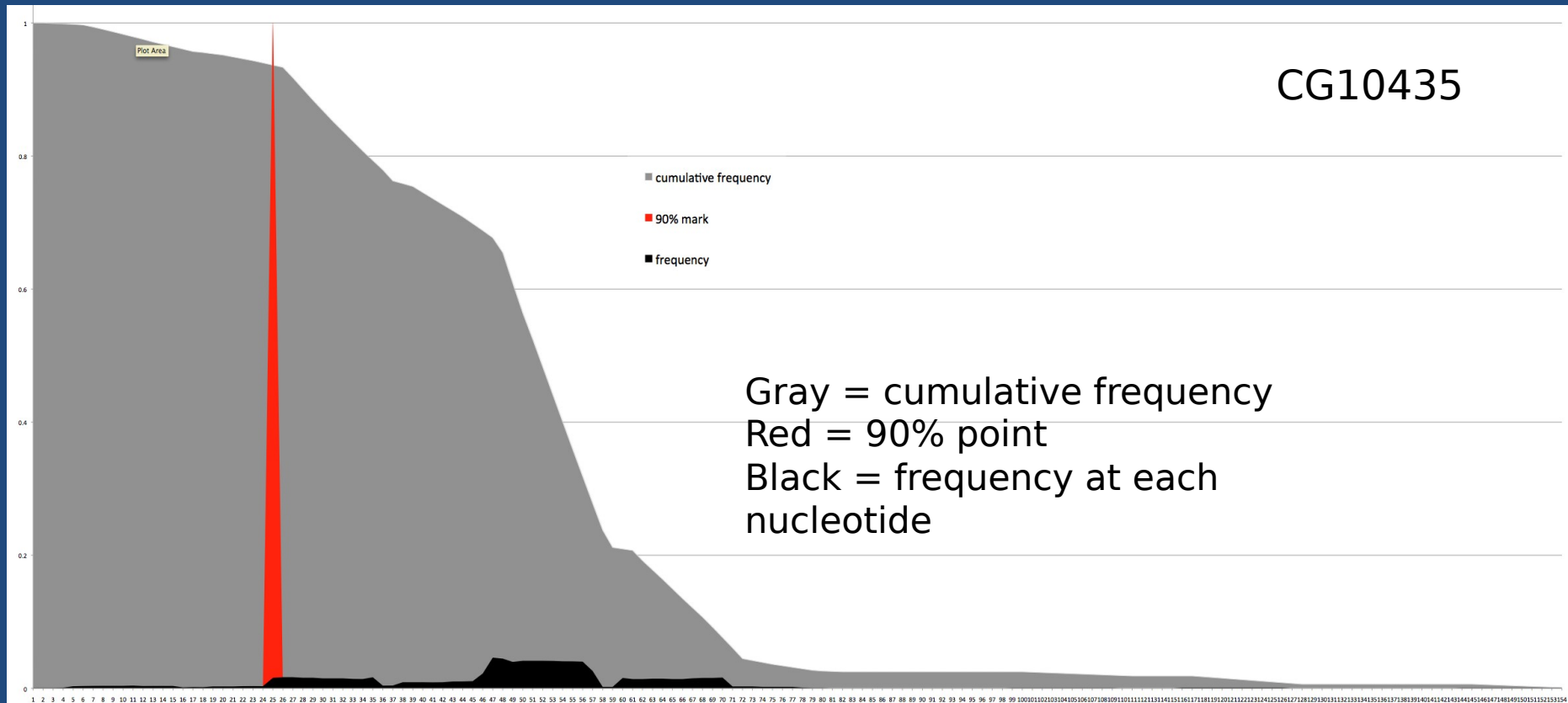


TSS annotation at 90% point peaked promoter class

TwdIL



TSS annotation at 90% point broad promoter class

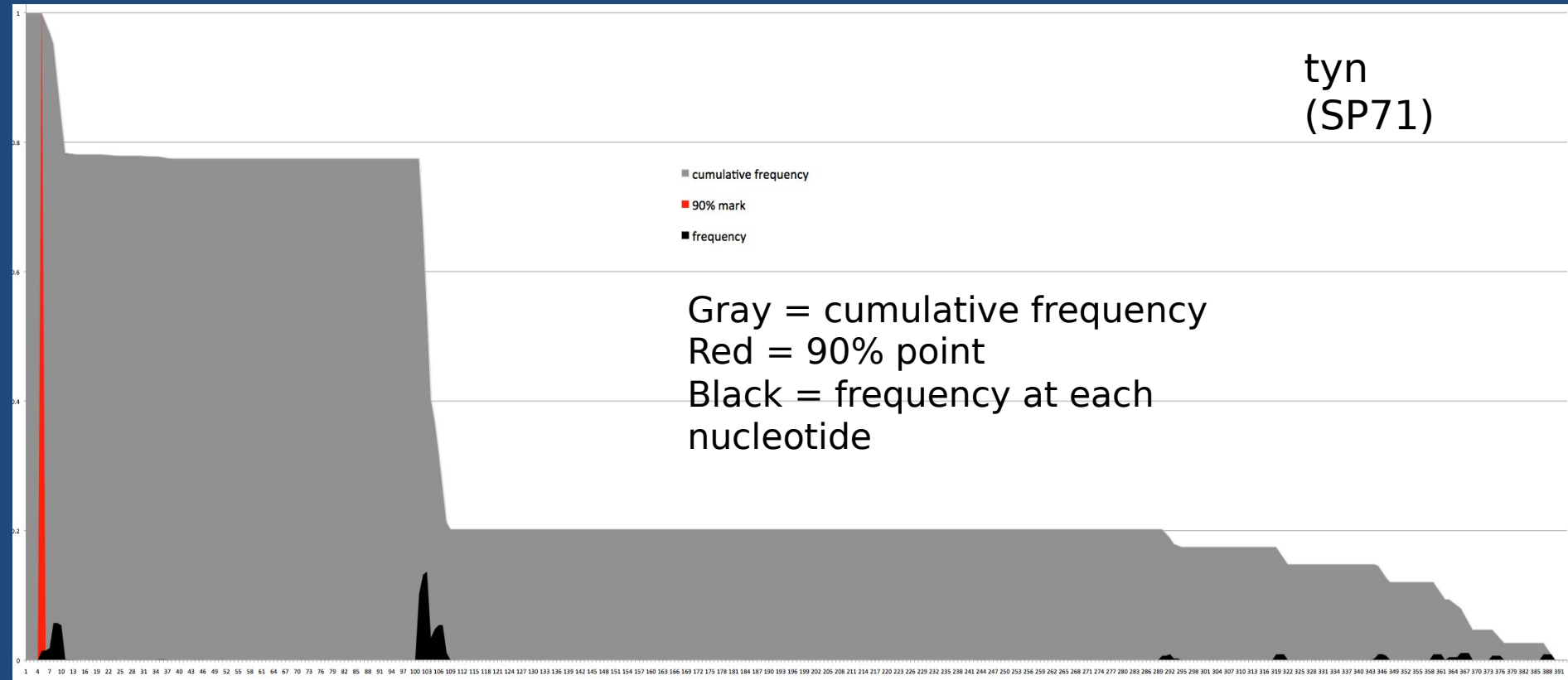


TSS annotation at 90% point broad promoter class

tyn
(SP71)

- cumulative frequency
- 90% mark
- frequency

Gray = cumulative frequency
Red = 90% point
Black = frequency at each
nucleotide



- Annotation on the genome of transcript features
 - RNA editing sites
 - Transcription start sites (as regions)
- Annotated transcripts will reflect TSS 90% point

New high-throughput data

- RNA-Seq junction calls
- Stranded RNA-Seq coverage data from multiple tissues (cell lines, treatments)
- Stop-codon readthrough predictions
- A-to-I RNA editing sites
- Transcription start site (TSS) mapping

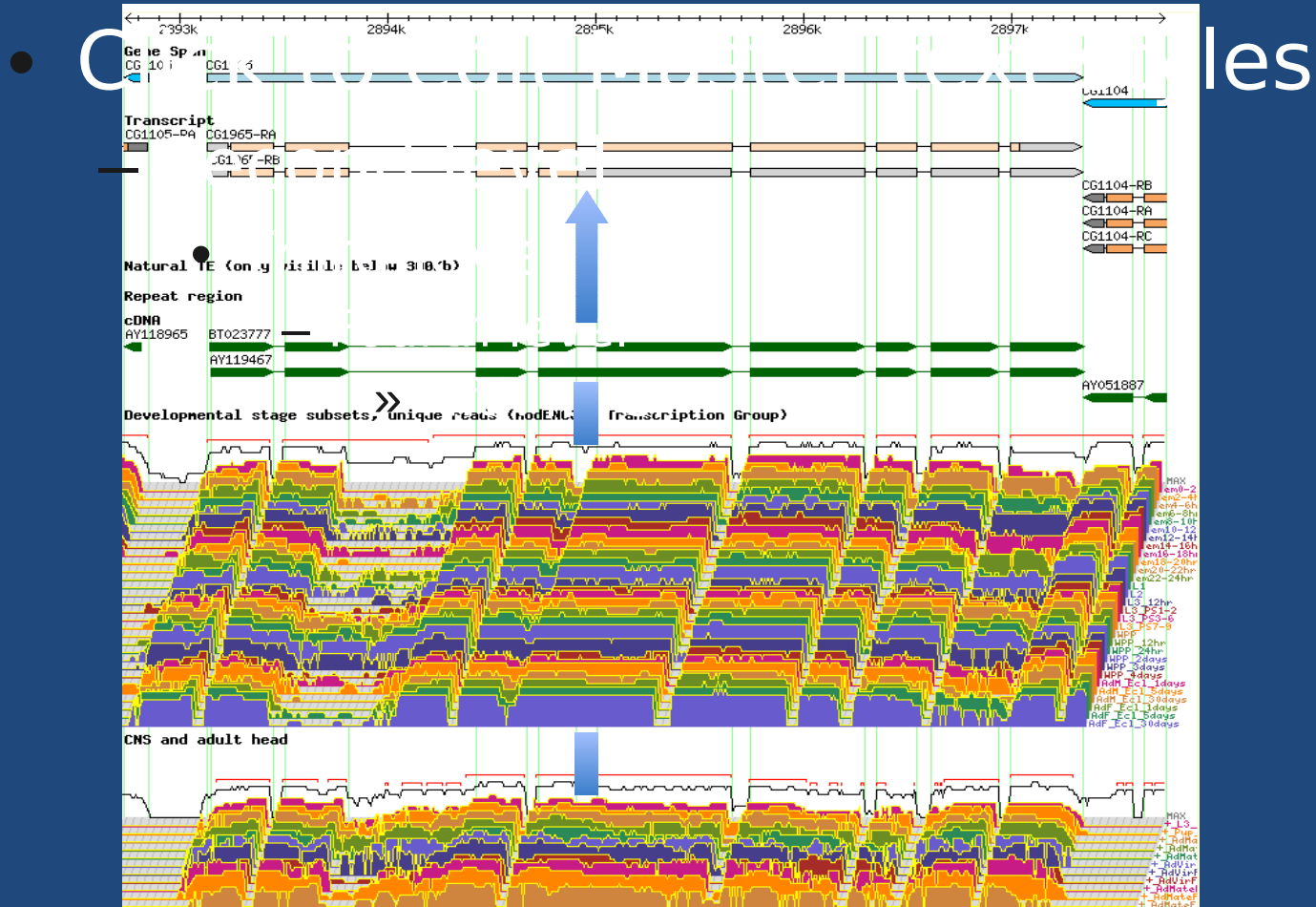
Integration of High Throughput Data

- Presentations in GBrowse
- Sequence feature reports
- New and improved annotations
- Assessing the data
- Annotation on the genome of transcript features
- Use of gene model comments

modENCODE papers

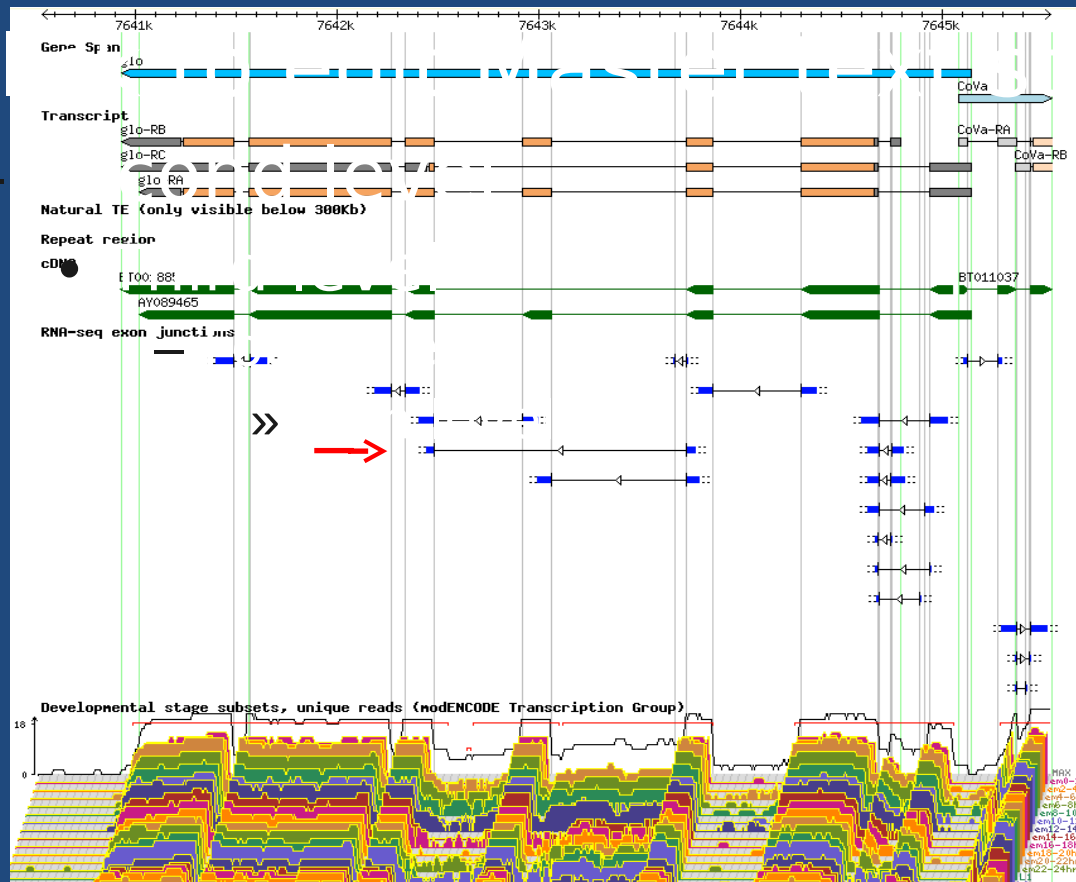
- RNA-Seq (coverage, junctions, 3' UTRs)
 - _ Graveley, et al. (2011) Nature 471:473-479
 - _ Cherbas, et al. (2011) Genome Res. 21:301-314
 - _ Smibert, et al. (2012) Cell Reports 23 Feb. 2012
- Stop-codon readthrough
 - _ Jungreis, et al. (2011) Genome Res. 21:2096-2113
- A-to-I RNA editing
 - _ Graveley, et al. (2011) Nature 471:473-479
- Transcription start site data
 - _ Hoskins, et al. (2011) Genome Res. 21:182-192

Transcripts that result in truncated polypeptides: retained introns



Transcripts that result in truncated polypeptides: alternative splices

- C



Annotated transcripts with truncated polypeptides (520 transcripts, 413 genes)

- Retained intron \Rightarrow premature stop
- Retained intron \Rightarrow downstream start
- Alternative splice \Rightarrow premature stop
- Alternative splice \Rightarrow downstream start

Comments (associated with transcript)

- Evidence supports alternative splice leading to premature stop codon and/or downstream start; may or may not produce functional polypeptide.
- Based on cDNA(s) with retained intron; results in premature stop codon and/or downstream start; may or may not produce functional polypeptide.

A few cases addressed in the literature

- Truncated polypeptides appear to be functional – can annotate properly
- The variants thought to be non-coding transcripts – cannot annotate properly
 - Change: create non-coding transcripts within a coding gene model
 - How do we handle the rest??

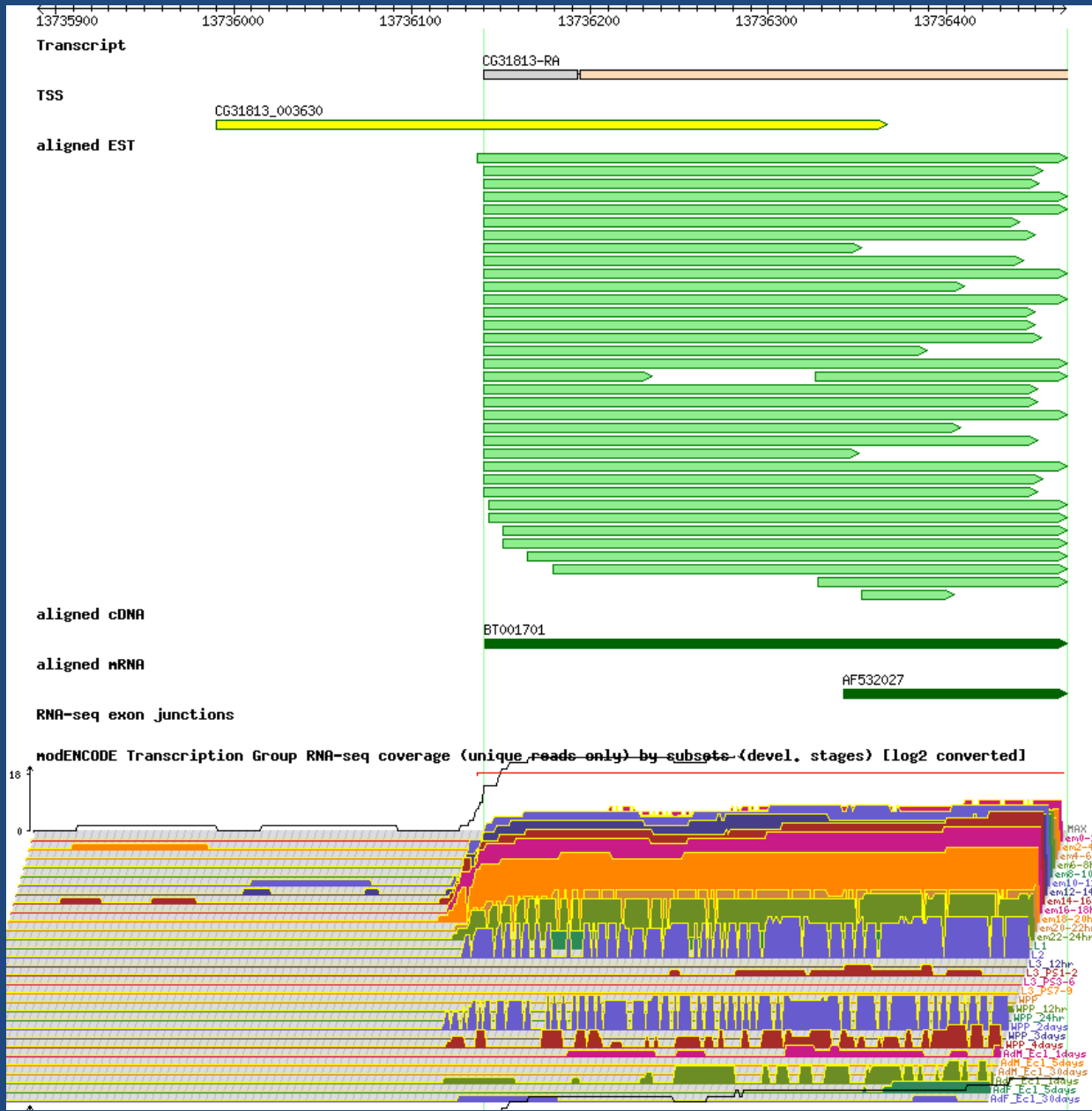
Some genes involved in RNA-splicing with annotated truncated isoforms

- tra
- su(w[a])
- Sxl
- B52
- Moca-cyp
- Rm62
- sqd
- CG1965
- CG1622
- CG1646
- Ars2
- Pep
- SC35
- scaf6
- U2af50
- Zn72D
- CG11266
- CG5316
- CG6686
- CG7757

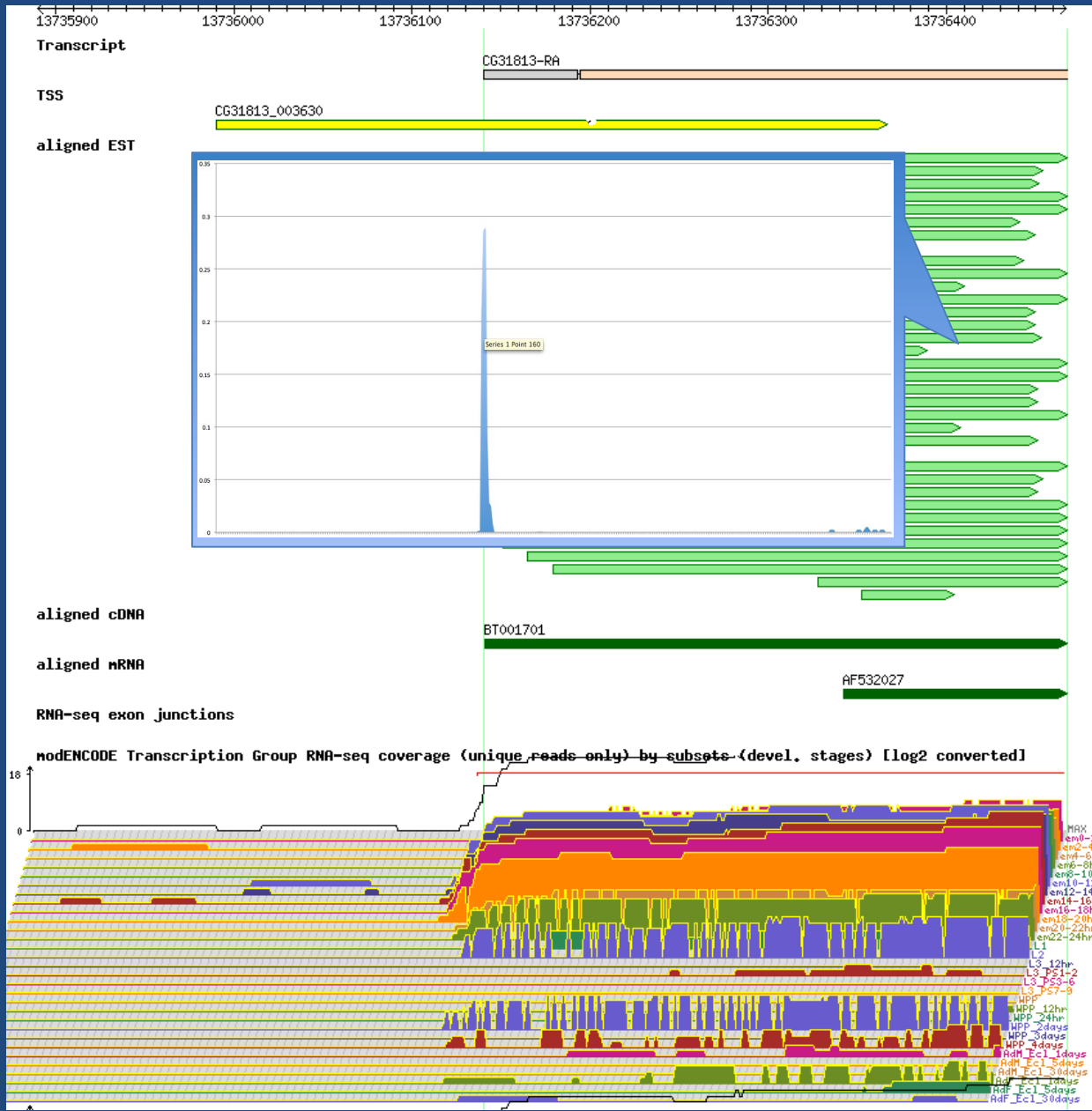
Short polypeptides supported in literature

- wah
- ade3
- svr
- CG344
22
- psq
- scf
- AGO
2
- osk

Gene annotation using modENCODE TSS



Gene annotation using modENCODE TSS



modENCODE TSS: FlyBase GBrowse

Instructions [\[Help\]](#) [\[Reset\]](#) [\[Bookmark this\]](#)

Search using a sequence name or range, gene symbol (but not full name, for example, ct but not cut), gene identifier, insertion symbol, insertion identifier, or other landmark. Searches are case-sensitive. Wildcards are not supported at this time. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: [cnn](#), [FBgn000490](#), [X:200000..220000](#), [2L:80,000..100,000](#), [2R:80,000..100,000](#), [3L:80,000..100,000](#), [3R:80,000..100,000](#), [4:100000..120000](#).

Search

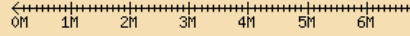
Landmark or Region:

X:17596976..17597975

Data Source

Overview

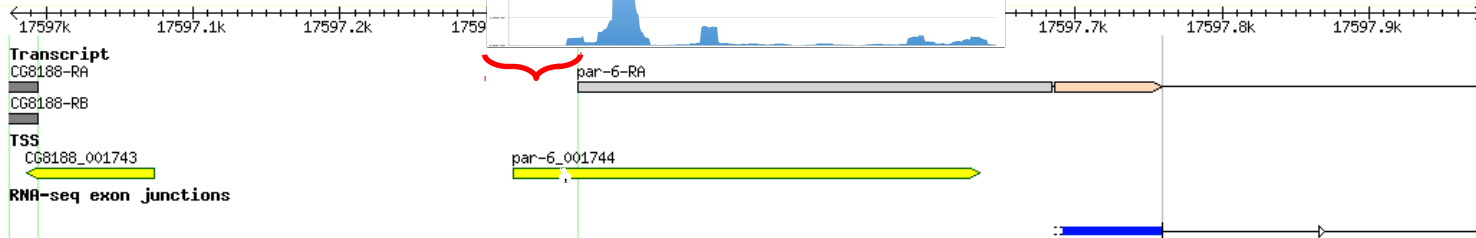
Overview of X



Report & Analysis tools:

Scroll/Zoom: Flip

Details



[Clear highlighting](#)

[\[Help and key for Stock Center Data evidence tiers\]](#)

Tracks

- Chip-seq data**
 - Chromatin Domains (5-state model, Kc cells)
 - Chromatin Domains (9-state model, S2 cells)
 - Chromatin Domains (9-state model, BG3 cells)
- Gene Model features**
 - Gene Span
 - Transcript
- Genome reagents and data**
 - aligned EST
 - aligned cDNA
 - aligned mRNA
- Genomic features**
 - Natural transposon
 - Transgene insertion site
 - Estimated Cytological band
- Mapped features**
 - Insulator class I
 - Insulator class II
- Other RNA-seq data**
 - cell lines expression (by strand) [log2 converted]
 - expression under treatments (by strand) [log2 converted]
- RNA-seq data**
 - RNA-seq exon junctions
 - modENCODE Transcription Group RNA-seq coverage (unique reads only) by subsets (devel. stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage (embryonic stages, by strand) [log2 converted]
 - K. White group RNA-seq coverage by subsets (devel.stages) [log2 converted]
 - Baylor group RNA-seq coverage by subsets (devel.stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage (all reads) by subsets (devel. stages) [log2 converted]
 - modENCODE Transcription Group RNA-seq coverage by subsets (tissue culture cells) [log2 converted]
- RNA-seq data, expression by tissue**
 - digestive system (by strand) [log2 converted]
 - imaginal disc and other carcass (by strand) [log2 converted]
 - gonads and male accessory glands (by strand) [log2 converted]
 - fat body and salivary glands (by strand) [log2 converted]
 - CNS and adult head (by strand) [log2 converted]
- External Annotation Tracks**

GBrowse: navigation options

- **Click to edit master text styles**

Second level

Examples: `cnn`, `FBgn0000490`, `X:200000..220000`, `2L:80,000..100,000`, `2R:80,000..100,000`, `3L`

Search **level** **Report & A** **Download**

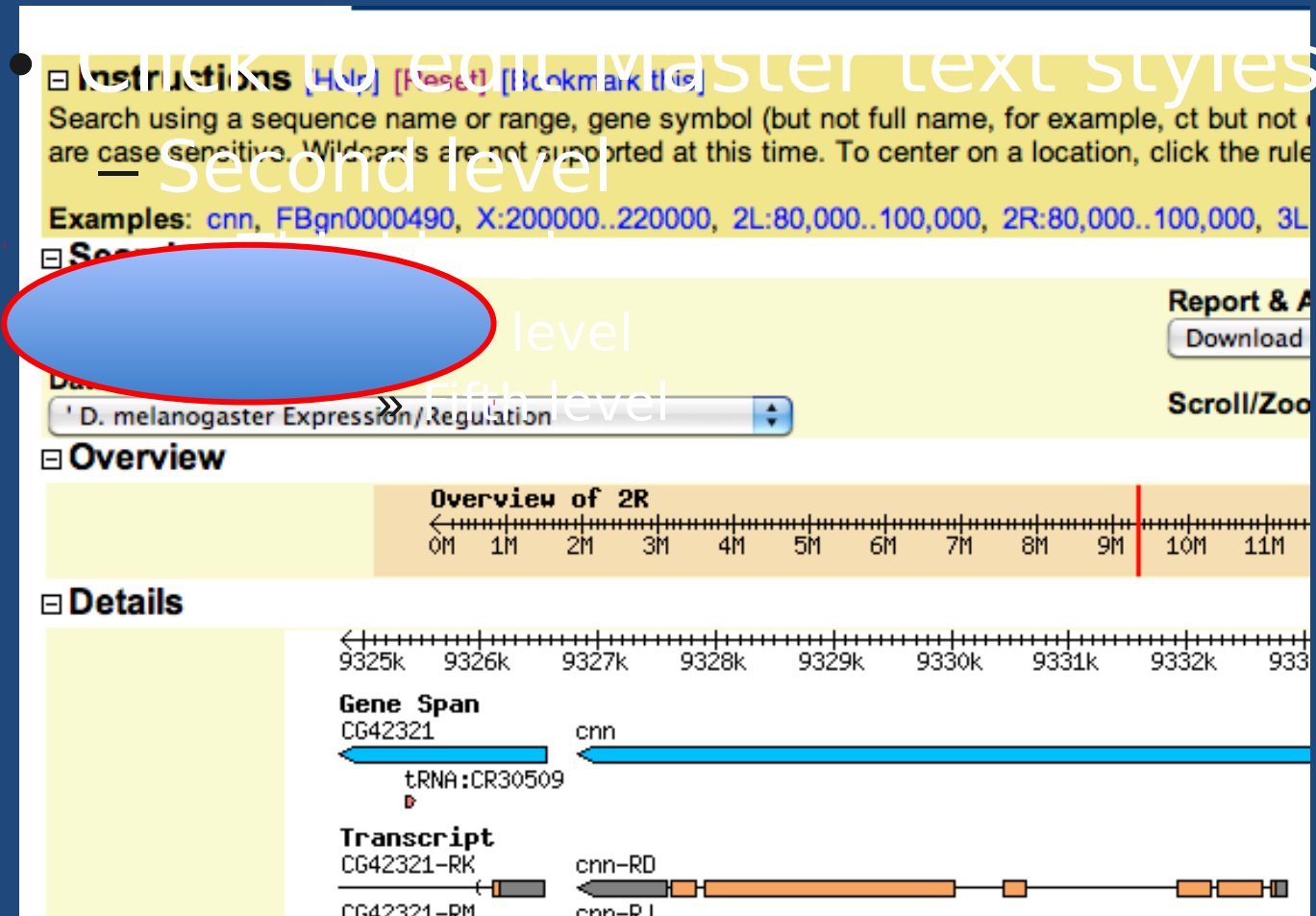
Overview **Overview of 2R** **Scroll/Zoom**

Details

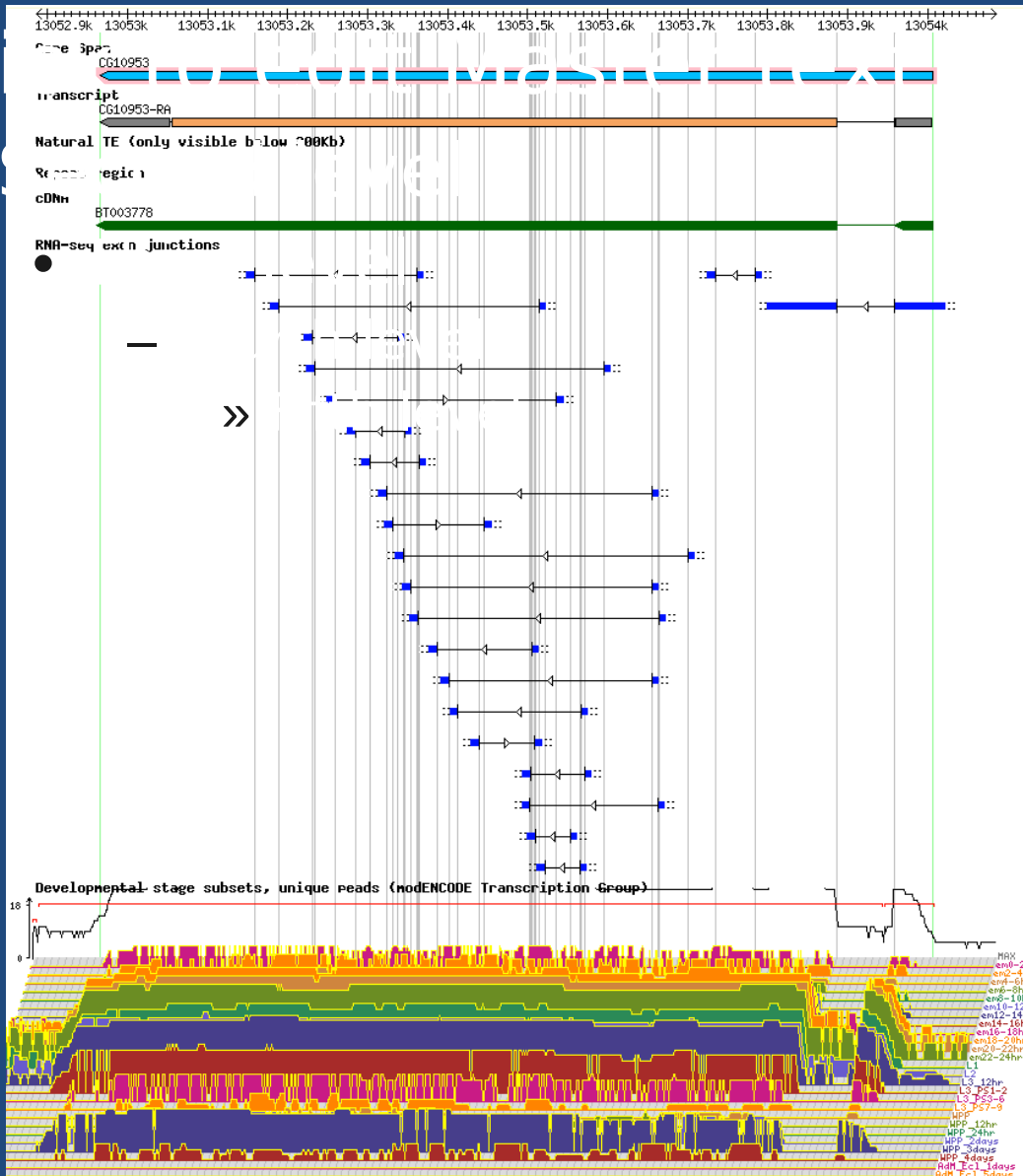
9325k 9326k 9327k 9328k 9329k 9330k 9331k 9332k 9333k

Gene Span
CG42321 `cnn`
tRNA:CR30509

Transcript
CG42321-RK `cnn-RD`
CG42321-RM `cnn-RJ`



Spurious junctions in region containing repeats



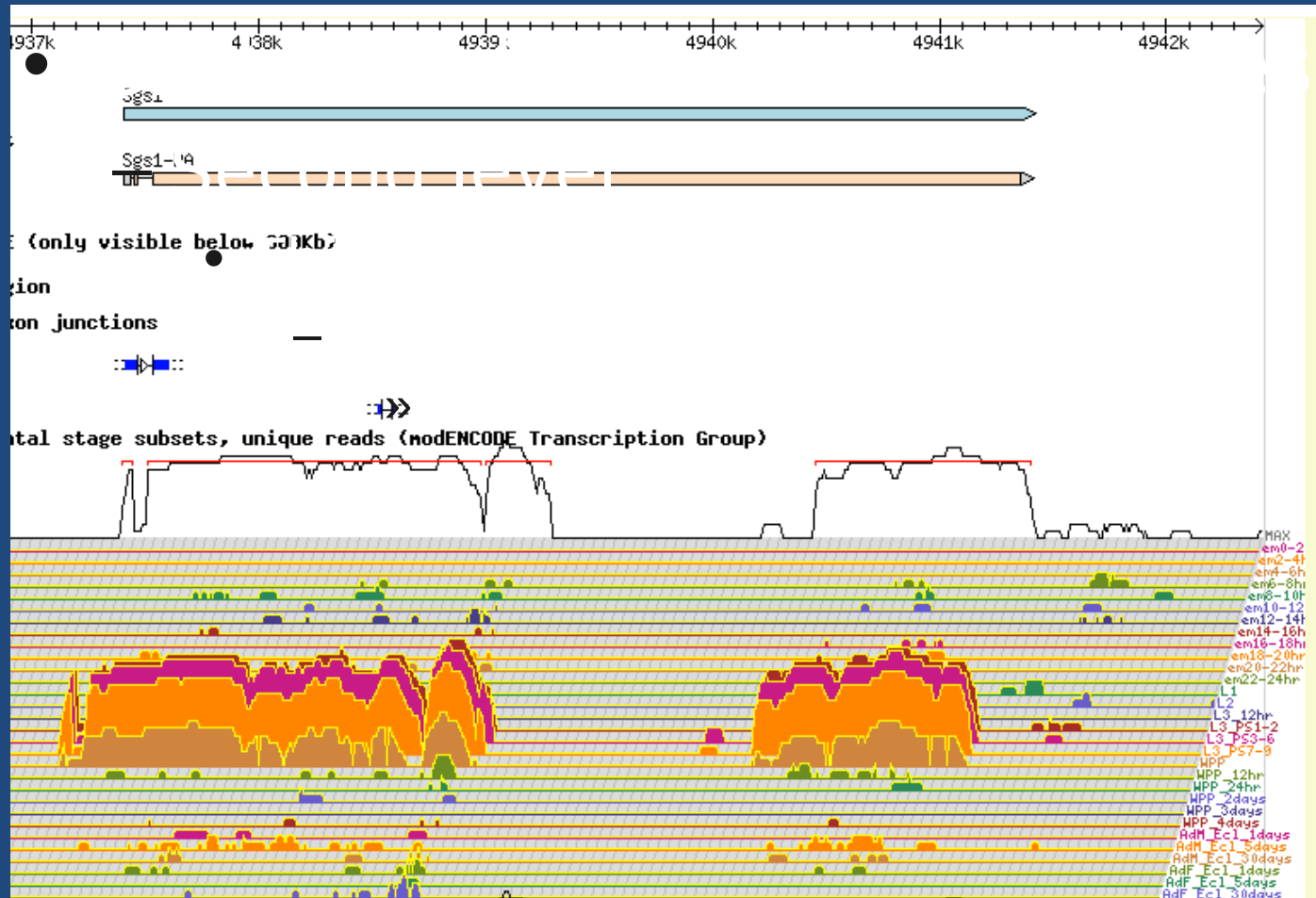
```

AGAAGCTACTCGATCGCTCTTGGAGTAACAGATTGCAGCTGTCATCATGA
AATTCTGGTTGTTGCCTTTGCCGTTCTGGCCTGCGCCTATGGTGATGTC
TCCCATTTGGGTTACGACTACTCTCCTCCGGCCCCAGCTCCGGTTTACCA
GCCCGCCCCGGCTCCGGTCTACCAGCCCGCCCCCGCTCCGGTCTACCAGC
CTGCTCCAGCACCAGTTGTGATCCCTGCTCCTGCTCCAGCTCCAGTTGTG
ATCCCTGCTCCTGCTCCCGTCCAGTTGTGATTCCGGCTCCAGCTCCCGT
GAAGACGTACGTCCACCCGCACCCATCAGCATCCCAGCCCCAGTTTACC
AGCCAGCTCCTGCTCCATTAGGATCCAGCCCCAGTTTACCAGCCAGCT
CCTGCTCCATCAGCATCCA [REDACTED] TGAGATCCCTGCC
AGCTCCAGTGAACACCTACATTCCCCCGCACCAGCACCAGCTCCGGTTT
ACCAGCCAGCCCCGCTCCATTCCCGTGAGCATCCCGGCTCCAGTCCC
GTTTACCAGCCTGCTCCCGTCCCGTTGTGATCCAGCCCCGCTCCAGC
ACCAGTTGTGATCCCTGCTCCTGCTCCCGTCCAGTTGTGATCCAGCCC
CAGCTCCGTTAAGTCTATGTGCCACCTGCACCAATCAGCATCCCAGCC
CCGGCTCCCGTTTACCAGCCAGTCCCATCAGCATCCCAGCCCCCTGCC
AGTTTACCAACCAGCTCCCGCTCCCGTCTACCAGCCAACCAACTCAGG
TTCTGGAGGAGATCGAGCCCGCTTCCAATGATGGATACCGCTACAAGACC
GTGCGTGTCCGCTACCGTCCCGTTTCTAAACTACCAAAGCCCAAC
GAAAACTGTTTCTAACGAATCCCTACGGGGACTCGCAATAAAAAATTT
CCGCTTCAATTTGTTAAAC
    
```

```

MKFLVFAVAVLACAYGDVSHLGYDYSPPAPAPVYQPAPAPVYQPAPAPVY
QPAPAPVVIAPAPAPVVIAPAPAPVVIAPAPAPVVIAPAPVVIAPAPV
YQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPV
VYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPV
APAPVKSIVYPPAPVYQPAPAPVYQPAPAPVYQPAPAPVYQPAPAPVY
QVLEIEPASNDGYRYKTVRRRVYRHRF
    
```

Repeat and low complexity regions create alignment problems



Sgs1 repeats foil RNA-Seq alignments

lick
Seq

- T

```
MIK) AL IFL QVSLI I QVKNVH AMY DWQV MQDQ PS EE LI PGCG GD CE NYTD  
PvQpCDTDSinPT I I KPRQKTKRPKS I RKT I KRI KRPRRKTTKW I KRATK  
RTTKRTRRRPTTPKTPDITDSDPITTTGAECTCSDRTTASSTDSTDRIT  
VTNTDWTI PLI LDTF PCTCE ESSTAIPSSPCIDTSTVIPTSPCTQETTT  
P I F T C S T Q G T Q T F C Y C A G T T T P R S T T T T S R P T T T T P R S T T T T T S R  
P T T T T P R S T T T T T R R P T T T T P R C T T T T S C A P T T T T P R S T T T T T S R P T  
T T P R C T T T T S T C S P T T T T P R S T T T T S R P T T T T P R C T T T T P S T S R P T T T  
T P R S T T K T S C A P T T T T P R T T T P S T S R P T T T T P R S T T T T S R P T T T T P  
R S T T T T T R R P T T T T P R S T T T T S R P T T T T P R S T T T T T S R P T T T T P R S  
T T T T C T C S P T T T T P R S T T T T S T S R P T T T T P R S T T T T S T S G P T T T T P R S T T  
T T T T S G P T T T T P R S T T T T C T C S P T T T T P R S T T T T P S T S R P T T T T P R S T T T T  
C T C S P T T T T P R S T T T T S T S R P T T T T P R S T T T T T S R P T T T T P R S T T T T S T  
S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G  
P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T  
T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T  
T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P  
R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S  
T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T T T S T S G P T T T T P R S T T  
T T S T S G P T T T T P R S T T T T S T C P T T T T P R S T T T T C T S G P T T T T P R S T T T T  
C T S C P T T T T P R S T T T T C T S C P T T T T P R S T T T T C T S G P T T T T P R S T T K T S T  
C A P T T T T P R S T T T T S T S R P T T T T P R S T T T T T S R P T T T T P R S T T T T P S T S R  
P T T T T P R S T T T T S T S R P T T T T P R S T T K T S C A P T T T T P R S T T T T T S R P T  
T T T P R S T T T T T S R P T T T T P R S T T T P C T S R P T T T T P R S T T T T T S R P T T T  
T P R S T T T T P C P T T T P S A S P T R T T T T R R P C P C H P Q P P Y Q I P P W S W W Y P N P T Y  
P N P V W P W Q P N P V V P Q W P Q M P G Y P N Q W P Q L P G Y P Q Q L P P P L P S P Q W P W S W P  
K P P V V Q P G C E N I C E N L L K G V K Y Q E D L I R R C L C T R T
```

styles

```
ATCCACAACAACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGAT  
CCACAACAACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCC  
ACAACAACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCAC  
AACAACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACA  
CAACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACA  
ACAACACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAAC  
AACTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACA  
CTTCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACT  
TCCACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTC  
CACAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCA  
CAAGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCACA  
AGTGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCACAAG  
TGGCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCACAAGTG  
GCCCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCACAAGTGGC  
CCAACGACAACACTACTCCTAGATCCACAACAACAACACTTCCACAAGTGGCC  
AACGACAACACTACTCCTAGATCCACAACAACAACACTTGCACAAGTGGCCCAA  
CGACAACACTACTCCTAGATCCACAACAACAACACTTGCACAAGTGGCCCAA  
ACAACACTACTCCTAGATCCACAACAACAACACTTGCACAAGTGGCCCAAACGAC  
AACTACTCCTAGATCCACAACAACAACACTTGCACAAGTGGCCCAAACGACAA  
CTACTCCTAGATCCACAACAACAACACTTGCACAAGTGGCCCAAACGACAAC  
ACTCCTAGATCCACAACAACAACACTTCCACAAGTCGCCCAACGACAACACTAC  
TCCTAGATCCACAACAACAACACTACCACAAGTCGCCCAACGACGACTACTC  
CTAGATCCACAACAACAACACTTCCACAAGTCGCCCAACGACAACACTACTCCT  
AGATCCACAACAACACTACTTCCACAAGTCGCCCAACGACAACACTACTCCTAG  
ATCCACAACAACAACACTTCCACTTGTGCCCAACGACAACACTACTCCTAGAT
```

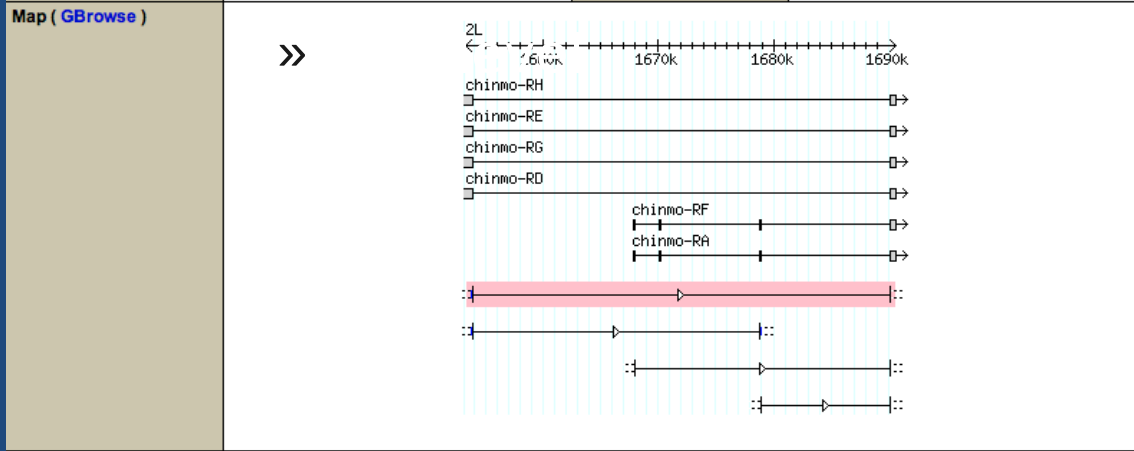
Click FlyBase Master text styles

Home Tools Files Species Documents Resources News Help Archives [Jump to Gene](#) [Go](#)

Profile Manager [+](#) [-](#) [?](#) [Help](#) [Open All](#) [Close All](#)

General information			
Symbol	Dmel:r5:2L:1653835:1690117:+	Species	<i>D. melanogaster</i>
Feature type	exon_junction	FlyBase ID	FBsf0000101966
Collection	modENCODE_mRNA-Seq_U_junctions BCM_1_RNAseq_junctions	Associated gene(s)	chinmo

Genomic Location	
Chromosome (arm)	2L
Sequence location	2L:1,653,835..1,690,117 [+]



- Recent Updates
- Sequence Data
- Associated Information
- Experimental Data
- RNA-Seq Read Counts
- Library: BCM_1_RNAseq_junctions
- Library: modENCODE_mRNA-Seq_U_junctions

modENCODE_mRNA-Seq_U_junctions	Expression stage(s)	Read Count
	embryonic stage 1 – 4	2
	embryonic stage 4 – 9	0
	embryonic stage 9 – 11	1
	embryonic stage 11 – 12	5
	embryonic stage 12 – 13	11
	embryonic stage 13 – 15	12
	embryonic stage 15 – 16	117