

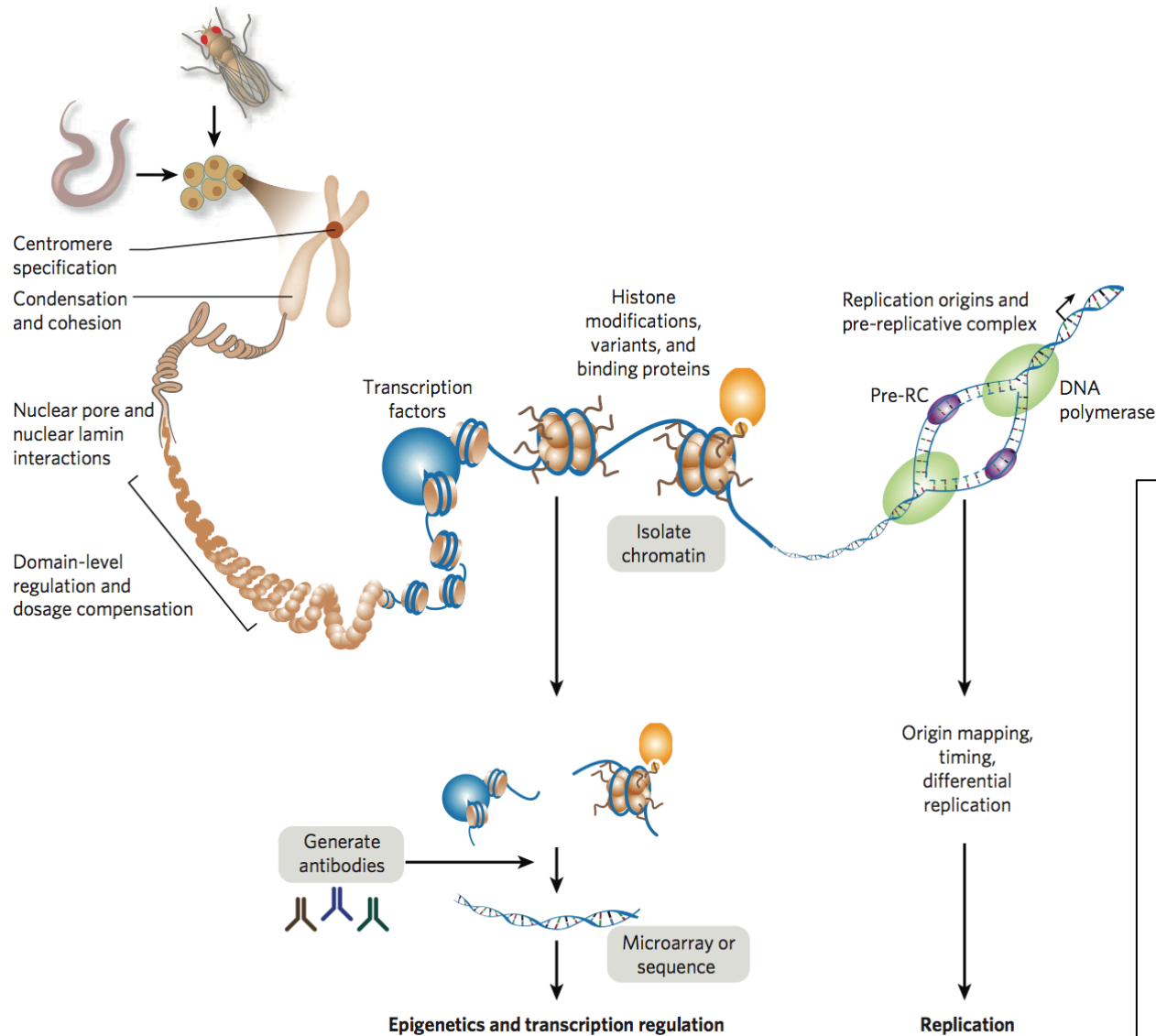


# Navigating High-throughput Data in FlyBase



# High-throughput data overview

(Celniker, 2009)



## ChIP

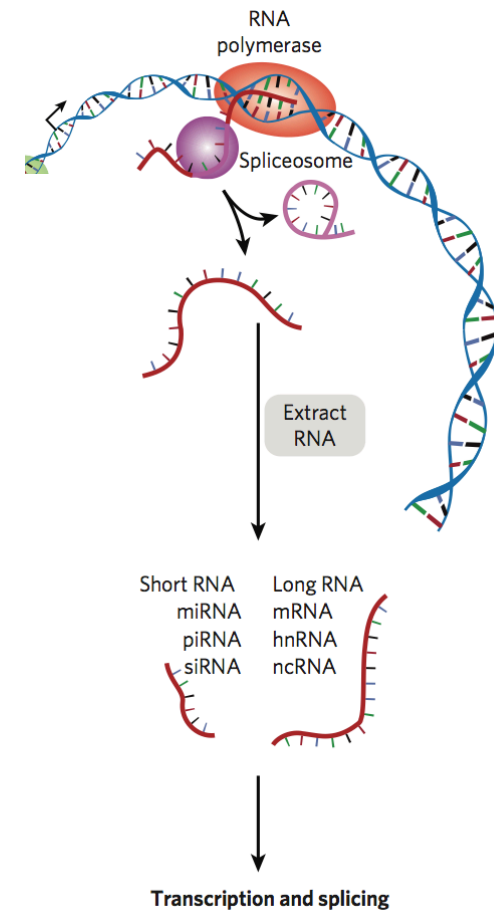
- > chromatin domains
- > transcription factor binding sites
- > insulators
- > enhancers
- > replication origins

# High-throughput data overview

(Celniker, 2009)

## RNA-Seq

- > transcriptome
- > exon junctions
- > A-I editing sites
- > transcription starts
- > new genes




# High-throughput Data: Views, reports, and queries

- GBrowse options (GBrowse 2)
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- Sequence feature reports
- FeatureMapper
- Querying RNA-Seq expression data

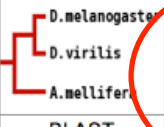
# High-throughput Data: Views, reports, and queries

- **GBrowse options (GBrowse 2)**
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- Sequence feature reports
- FeatureMapper
- Querying RNA-Seq expression data

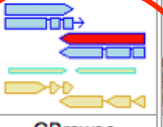
# High-throughput data in GBrowse: getting there


FB2014\_01, released January 17th, 2014  
 A Database of *Drosophila* Genes & Genomes


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



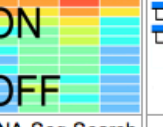
BLAST



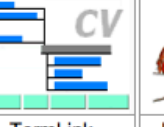
GBrowse




QueryBuilder




RNA-Seq Search



TermLink



ImageBrowse



Batch Download

---

**Fast-Track Your Paper**

[FlyBase Forum](#)

[Find a Fly Person](#)

**News**

[New in Release FB2014\\_01](#) | 17 Jan 14

[2014 Release Schedule](#) | 15 Jan 14

[FlyBase 102](#) | 1 Dec 13

[Stock Center User Survey](#) | 18 Jun 13

[Fly Mating Scheme Design](#) | 13 Mar 13

[GenBank Release](#) | 25 Jan 13

[NAR article on bibliography](#) | 5 Nov 12

[White Paper 2012](#) | 8 Jun 12

**Upcoming Meetings**

[55th ADRC](#) | 26 Mar 14

[Reg Dros Meeting 2014](#) | 27 May 14

[19th Crete Dros Mol/Dev Bio](#) | 22 Jun 14

[7th Intl Symp Mol Insect Sci](#) | 13 Jul 14

[XV Euro Dros Neurobiology](#) | 5 Oct 14

[Dros in Exp Genetics & Bio](#) | 6 Oct 14

[2nd ESF-EMBO Minibrains](#) | 8 Nov 14

**Courses**

[Dros Species Workshop XII](#) | 25 Mar 14

[W-T Dros Genetics/omics](#) | 27 Jul 14

**QuickSearch** ?

[Simple](#) [Expression](#) [Phenotype](#) [GO](#) [References](#) [Data Class](#)

Species:  include non-Dmel species

Enter text:

Note: Wild cards (\*) can be added to your search term

**Commentary** [See all commentaries](#)

**FlyBase RNA-Seq RPKM data bulk download**

Sample	RPKM Count
digestive system, 1-day adult	2
digestive system, 4-day adult	4
digestive system, 20-day adult	3
fat body, larvae L3 wandering	4
fat body, white prepupae	40
fat body, pupae P8	25
carcass, larvae L3 wandering	15
carcass, 1-day adult	15
carcass, 4-day adult	16
carcass, 20-day adult	11
ovary, virgin 4-day female	66
ovary, mated 4-day female	54
testis, mated 4-day male	65




*May 6, 2013.* FlyBase is extending its initial gene-level analyses of RNA-seq throughput data from modENCODE and others. The algorithm for RPKM (reads per kilobase per million mapped reads) has been refined, additional datasets have been analyzed, and these data are now available for bulk download... [\(More\)](#)

---

FlyBase is supported by a grant from the [National Human Genome Research Institute](#) at the U.S. [National Institutes of Health](#) #P41 HG000739. Support is also provided by the [British Medical Research Council](#), the [Indiana Genomics Initiative](#), and the [National Science Foundation](#) through XSEDE resources provided by [Indiana University](#). [Copyright Statement](#).

version FB2014\_01, released January 17, 2014

[Contact FlyBase](#)   [Cite FlyBase](#)   [Site Map](#)



# Navigating GBrowse 2: Important new features

The screenshot displays the GBrowse 2 interface with several key features highlighted by red circles:

- Select Tracks:** Located in the top navigation bar, this link allows users to manage the visibility of different data tracks.
- Clear highlighting:** Located at the bottom of the interface, this link is used to reset the visual highlighting on the tracks.

The main interface includes a search bar, a "Landmark or Region" input field with a "Search" button, and a "Data Source" dropdown menu. The "Overview" section shows a genomic map with a scale from 0M to 20M. The "Details" section provides a zoomed-in view of a 2 kbp region, showing various tracks such as Gene Span, Transcript, Transgenic Insertion Site, Natural TE, Repeat region, cDNA, RNA-seq exon junctions, and Developmental stage subsets. The bottom of the interface features a "CONGO" track and a "Select Tracks" button.

- Data Source menu – new 'All Features' view
- Easy to delete or hide track data
- Tracks are moveable – up and down
- Links to information/metadata for tracks (here and on Tracks page)
- Select Tracks – two links.



# Navigating GBrowse 2: Track Selection

The screenshot shows the 'Select Tracks' interface in GBrowse 2. The 'Browser' tab is circled in red. The 'Noncoding Features' track is selected and expanded, showing a grid of sub-tracks. Red arrows point to the 'Noncoding Features' header and the 'TFBS - zinc finger domain' track.

**Browser** Select Tracks Custom Tracks Preferences

Back to Browser Show Favorites Only Clear All Favorites

Tracks

- Reference Genome Annotations (Iso-1)  All on  All off
  - Gene Span [?]
  - Transcript [?]
  - CDS [?]
  - Natural TE (only visible below 300Kb) [?]
  - Repeat region [?]
- General  All on  All off
  - Estimated Cytological Band [?]
  - 3-frame translation (forward) [?]
  - DNA/GC Content [?]
  - 3-frame translation (reverse) [?]
- Aligned Evidence  All on  All off
  - cDNA [?]
  - EST [?]
  - mRNA [?]
  - other aligned sequences [?]
  - RNA-seq exon junctions [?]
  - PeptideAtlas peptides [?]
  - Transcription Start Sites (embryonic) [?]
- Previous Gene Models  All on  All off
  - Recently deleted transcripts [?]
- Mapped Mutations  All on  All off
  - Transgenic Insertion Site [?]
  - Point Mutation [?]
  - Sequence Variant [?]
  - Uncharacterized Change in Nucleotide Sequence [?]
  - Aberration Junction [?]
  - Complex Substitution [?]
  - Indels [?]
  - Rescue Fragment [?]
- Gene Predictions  All on  All off
  - Augustus [?]
  - Genscan [?]
  - Genie [?]
  - GeneID [?]
  - MCB1 Gnomon [?]
  - CONTRAST [?]
- Similarity
  - Synteny features  All on  All off
    - Orthologs [?]
  - Proteins  All on  All off
    - D. melanogaster proteins [?]
    - Other proteins [?]
- Noncoding Features**  All on  All off
  - 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 (cells) [?]
  - Chromatin Domains (cells) [?]
  - Chromatin Domains (cells) [?]
  - Origin of Replication
  - Putative Brain Enhancer [?]
  - RNA Editing Sites [?]
- Microarray Features  All on  All off
  - Affymetrix v1 [?]
  - Affymetrix v2 [?]

# GBrowse tracks documentation

**Noncoding Features**  All on  All off

<input type="checkbox"/> Insulator class I <a href="#">[?]</a>	<input type="checkbox"/> TFBS - HOT spot analysis <a href="#">[?]</a>	<input type="checkbox"/> Chromatin Domains (5-state model, Kc cells) <a href="#">[?]</a>
<input type="checkbox"/> Insulator class II <a href="#">[?]</a>	<input type="checkbox"/> TFBS - zinc finger domain <a href="#">[?]</a>	<input type="checkbox"/> Chromatin Domains (9-state model, S2 cells) <a href="#">[?]</a>
<input type="checkbox"/> Protein Binding Site <a href="#">[?]</a>	<input type="checkbox"/> TFBS - homeodomain <a href="#">[?]</a>	<input type="checkbox"/> Chromatin Domains (9-state model, BG3 cells) <a href="#">[?]</a>
<input type="checkbox"/> Enhancer <a href="#">[?]</a>	<input type="checkbox"/> TFBS - helix-loop-helix domain <a href="#">[?]</a>	<input type="checkbox"/> Origin of Replication <a href="#">[?]</a>
<input type="checkbox"/> Silencer <a href="#">[?]</a>	<input type="checkbox"/> TFBS - BTB/POZ domain <a href="#">[?]</a>	<input type="checkbox"/> RNA Editing Sites <a href="#">[?]</a>
<input type="checkbox"/> Regulatory Region <a href="#">[?]</a>	<input type="checkbox"/> TFBS - other <a href="#">[?]</a>	

**Microarray Features**  All on  All off

<input type="checkbox"/> Affymetrix v1 <a href="#">[?]</a>	<input type="checkbox"/> Affymetrix v2 <a href="#">[?]</a>	
--	--	--

**Expression Levels**

**MicroArrays**  All on  All off

<input type="checkbox"/> Timecourse [cDNA] (Arbeitman et al.) <a href="#">[?]</a>	<input type="checkbox"/> Timecourse [amplicons] (Gauhar et al.) <a href="#">[?]</a>	<input type="checkbox"/> Tissue Expression [Affy2] (FlyAtlas) <a href="#">[?]</a>
---	---	---

**Expression Levels**

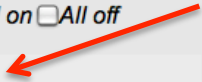
**RNA-seq**  All on  All off

<input type="checkbox"/> Developmental stage subsets (Baylor) <a href="#">[?]</a>	<input type="checkbox"/> Tissue culture cells (modENCODE Transcription Group) <a href="#">[?]</a>	<input type="checkbox"/> Treatments/Conditions, by strand <a href="#">[?]</a>
<input checked="" type="checkbox"/> Developmental stage subsets, unique reads (modENCODE Transcription Group) <a href="#">[?]</a>	<input type="checkbox"/> Tissue culture cells, by strand (modENCODE Transcription Group) <a href="#">[?]</a>	

**Expression Levels**

**RNA-Seq by Tissue**  All on  All off

<input type="checkbox"/> Digestive system <a href="#">[?]</a>	<input type="checkbox"/> CNS and adult head <a href="#">[?]</a>	<input type="checkbox"/> L3 CNS neuroblast <a href="#">[?]</a>
<input type="checkbox"/> Fat body and salivary glands <a href="#">[?]</a>	<input type="checkbox"/> Gonads and male accessory glands <a href="#">[?]</a>	
<input type="checkbox"/> Imaginal disc and other carcass <a href="#">[?]</a>	<input type="checkbox"/> L3 CNS neuron <a href="#">[?]</a>	



# GBrowse tracks documentation: tracks = datasets

**Expression Levels: RNA-Seq**

**Developmental stage subsets (Baylor) [FBic0000060](#)**

Available in these data source views:

- Expression/Regulation

**Developmental stage subsets, unique reads (modENCODE) [FBic0000085](#)**

Available in these data source views:

- Expression/Regulation

**Tissue culture cells (modENCODE Transcription Group) [FBic0000116](#)**

Available in these data source views:

- Expression/Regulation

**Tissue culture cells, by strand (modENCODE Transcription Group) [FBic0000260](#)**

Available in these data source views:

- Expression/Regulation

**Treatments/Conditions [FBic0000236](#)**

Available in these data source views:

- Expression/Regulation

---

**Expression Levels: RNA-Seq by Tissue**

**Digestive system** Comprised of four RNA-Seq data sets: [mE\\_mRNA\\_L3\\_Wand\\_dig\\_sys \(FBic0000227\)](#) [mE\\_mRNA\\_A\\_1d\\_dig\\_sys \(FBic0000219\)](#) [mE\\_mRNA\\_A\\_4d\\_dig\\_sys \(FBic0000223\)](#) [mE\\_mRNA\\_A\\_20d\\_dig\\_sys \(FBic0000221\)](#)

Available in these data source views:

- Expression/Regulation

**Fat body and salivary glands** [mE\\_mRNA\\_L3\\_Wand\\_fat \(FBic0000228\)](#) [mE\\_mRNA\\_WPP\\_fat \(FBic0000233\)](#) [mE\\_mRNA\\_P8\\_fat \(FBic0000235\)](#) [mE\\_mRNA\\_L3\\_Wand\\_saliv \(FBic0000230\)](#) [mE\\_mRNA\\_WPP\\_saliv \(FBic0000234\)](#)

Available in these data source views:

- Expression/Regulation

**Imaginal disc and other carcass** [mE\\_mRNA\\_L3\\_Wand\\_imag\\_disc \(FBic0000229\)](#) [mE\\_mRNA\\_L3\\_Wand\\_carcass \(FBic0000226\)](#) [mE\\_mRNA\\_A\\_1d\\_carcass \(FBic0000218\)](#) [mE\\_mRNA\\_A\\_4d\\_carcass \(FBic0000222\)](#) [mE\\_mRNA\\_A\\_20d\\_carcass \(FBic0000220\)](#)

Available in these data source views:

- Expression/Regulation

**CNS and adult head** [mE\\_mRNA\\_L3\\_CNS \(FBic0000225\)](#) [mE\\_mRNA\\_P8\\_CNS \(FBic0000224\)](#) [mE\\_mRNA\\_A\\_MateM\\_1d\\_head \(FBic0000209\)](#) [mE\\_mRNA\\_A\\_MateM\\_4d\\_head \(FBic0000216\)](#) [mE\\_mRNA\\_A\\_MateM\\_20d\\_head \(FBic0000214\)](#) [mE\\_mRNA\\_A\\_VirF\\_1d\\_head \(FBic0000210\)](#) [mE\\_mRNA\\_A\\_VirF\\_4d\\_head \(FBic0000211\)](#) [mE\\_mRNA\\_A\\_VirF\\_20d\\_head \(FBic0000231\)](#) [mE\\_mRNA\\_A\\_MateF\\_1d\\_head \(FBic0000207\)](#) [mE\\_mRNA\\_A\\_MateF\\_4d\\_head \(FBic0000213\)](#) [mE\\_mRNA\\_A\\_MateF\\_20d\\_head \(FBic0000212\)](#)

Available in these data source views:

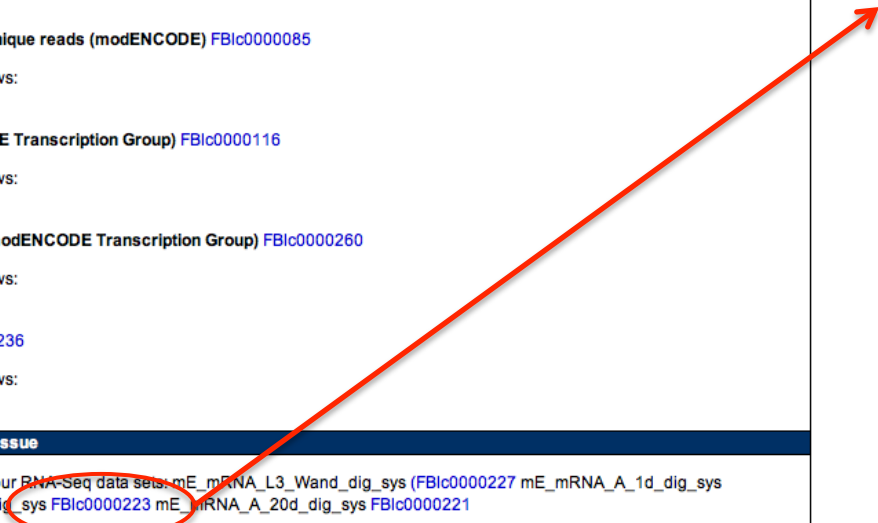
- Expression/Regulation

**Gonads and male accessory glands** [mE\\_mRNA\\_A\\_MateM\\_4d\\_testis \(FBic0000217\)](#) [mE\\_mRNA\\_A\\_MateM\\_4d\\_acc\\_gland \(FBic0000215\)](#) [mE\\_mRNA\\_A\\_VirF\\_4d\\_ovary \(FBic0000232\)](#) [mE\\_mRNA\\_A\\_MateF\\_4d\\_ovary \(FBic0000208\)](#)

Available in these data source views:

- Expression/Regulation

To Dataset report



# High-throughput Data: Views, reports, and queries

- GBrowse options (GBrowse 2)
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- Sequence feature reports
- FeatureMapper
- Querying RNA-Seq expression data

# Dataset reports

FlyBase  
 Dataset mE\_mRNA\_A\_4d\_dig\_sys  
 FB2014\_01, released January 17th, 2014

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

Help Open All Close All

General Information			
Name	mE_mRNA_A_4d_dig_sys	Species	<i>D. melanogaster</i>
Dataset type	RNA-Seq transcriptome coverage	FlyBase ID	FBic0000223
Source and Content			
Consists of	cDNA short sequencing reads from high-throughput sequencing. <a href="#">(Graveley et al., 2011.4.13)</a>		
Created by			
Available from			
Strain	Oregon-R-modENCODE <a href="#">(Graveley et al., 2011.4.13)</a>		
Stage and tissue			
Stage	Tissue/Position (including subcellular localization)	Reference	
day 4 of adulthood	pupal/adult digestive system	<a href="#">(Graveley et al., 2011.4.13)</a>	
Cell Line			
Recent Updates			
Description and Members			
Description	Short sequencing reads of cDNA prepared from polyadenylated RNA isolated from mixed mixed males/females, eclosion + 4 days, digestive system. Transcriptome represented as frequency of strand-specific reads along genome (see GBrowse presentation; data in wiggle format); this collection is composed of uniquely aligned reads. <a href="#">(Graveley et al., 2011.4.13)</a>		
Parent collections	<a href="#">modENCODE_mRNA-Seq_tissues</a>		
Component collection(s)			
Number in collection	140,809,354 <a href="#">(Graveley et al., 2011.4.13)</a>		
Comment on number in collection	Number of unique reads. <a href="#">(Graveley et al., 2011.4.13)</a>		
Members			
Experimental protocol			
Vector			
Sample preparation	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Collection preparation	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Mode of assay	Protocol: strand-specific Read length (bases):76 See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Assay platform	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Data analysis	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Additional data			
	More information is available under: <a href="#">modMine: modENCODE_3207</a> <a href="#">modMine: modENCODE_3454</a>		
Associated files			
Additional sites			
Synonyms and Secondary IDs			
Reported As			
Symbol Synonym	mE_mRNA_A_4d_dig_sys MixMF_4d_dig_sys <a href="#">(Graveley et al., 2011.4.13)</a>		
Secondary FlyBase IDs			
References ( 1 )			
Personal communication to FlyBase	Graveley et al., 2011.4.13, The <i>D. melanogaster</i> transcriptome: modENCODE RNA-Seq data for dissected tissues The <i>D. melanogaster</i> transcriptome: modENCODE RNA-Seq data for dissected tissues [FBrc0213503]		

strain →

stage & tissue →

description →

methods {

links to data repositories {

links to papers and supplementary files {

# Dataset reports – links to parent report

FB2014\_01, released January 17th, 2014

**FlyBase** Dataset mE\_mRNA\_A\_4d\_dig\_sys

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

Help Open All Close All

General Information			
Name	mE_mRNA_A_4d_dig_sys	Species	<i>D. melanogaster</i>
Dataset type	RNA-Seq transcriptome coverage	FlyBase ID	FBic0000223
Source and Content			
Consists of	cDNA short sequencing reads from high-throughput sequencing. <a href="#">(Graveley et al., 2011.4.13)</a>		
Created by			
Available from			
Strain	Oregon-R-modENCODE <a href="#">(Graveley et al., 2011.4.13)</a>		
Stage and tissue			
Stage	Tissue/Position (including subcellular localization)	Reference	
day 4 of adulthood	pupal/adult digestive system	<a href="#">(Graveley et al., 2011.4.13)</a>	
Cell Line			
Recent Updates			
Description and Members			
Description	Short sequencing reads of cDNA prepared from polyadenylated RNA isolated from mixed mixed males/females, eclosion + 4 days, digestive system. Transcriptome represented as frequency of strand-specific reads along genome (see GBrowse presentation; data in wiggle format); this collection is composed of uniquely aligned reads. <a href="#">(Graveley et al., 2011.4.13)</a>		
Parent collections	<a href="#">modENCODE_mRNA-Seq_tissues</a>		
Component collection(s)			
Number in collection	140,809,354 <a href="#">(Graveley et al., 2011.4.13)</a>		
Comment on number in collection	Number of unique reads. <a href="#">(Graveley et al., 2011.4.13)</a>		
Members			
Experimental protocol			
Vector			
Sample preparation	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Collection preparation	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Mode of assay	Protocol: strand-specific Read length (bases):76 See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Assay platform	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Data analysis	See report for <a href="#">modENCODE_mRNA-Seq_tissues</a> . <a href="#">(Graveley et al., 2011.4.13)</a>		
Additional data			
	More information is available under: <a href="#">modMine: modENCODE_3207</a> <a href="#">modMine: modENCODE_3454</a>		
Associated files			
Additional sites			
Synonyms and Secondary IDs			
Reported As			
Symbol Synonym	mE_mRNA_A_4d_dig_sys MixMF_4d_dig_sys <a href="#">(Graveley et al., 2011.4.13)</a>		
Secondary FlyBase IDs			
References ( 1 )			
Personal communication to FlyBase	Graveley et al., 2011.4.13, The <i>D. melanogaster</i> transcriptome: modENCODE RNA-Seq data for dissected tissues The <i>D. melanogaster</i> transcriptome: modENCODE RNA-Seq data for dissected tissues [FBrt0213503]		

strain →

stage & tissue →

description →

methods

links to data  
repositories

links to papers and  
supplementary files

# Dataset reports – Parent report


FB2014\_01, released January 17th, 2014  
**Dataset modENCODE\_mRNA-Seq\_tissues**

[Home](#) [Tools](#) [Files](#) [Species](#) [Documents](#) [Resources](#) [News](#) [Help](#) [Archives](#)

General Information		
Name	modENCODE_mRNA-Seq_tissues	Species: <i>D. melanogaster</i>
Dataset type	RNA-Seq transcriptome profiles	FlyBase ID: FBic0000206
Source and Content		
Consists of	cDNA short sequencing reads from high-throughput sequencing. <a href="#">(Graveley et al., 2011.4.13)</a>	
Created by		
Available from		
Strain	Oregon-R-modENCODE <a href="#">(Graveley et al., 2011.4.13)</a>	
Stage and tissue		
Cell Line		
<input type="checkbox"/> Recent Updates <input checked="" type="checkbox"/> Description and Members		
Description	Umbrella record for 29 collections that differ by dissected tissues from Oregon-R animals. Consists of 76-100nt strand-specific paired-end reads; longer reads trimmed so that all aligned reads are 76nt. Transcriptome represented as frequency of strand-specific reads along genome (data in wiggle format; see GBrowse presentation). The profiles are represented as uniquely aligned reads. <a href="#">(Graveley et al., 2011.4.13)</a>	
Parent collections		
Component collection(s)	<a href="#">mE_mRNA_A_MateF_1d_head</a> <a href="#">mE_mRNA_A_MateF_4d_ovary</a> <a href="#">mE_mRNA_A_MateM_1d_head</a> <a href="#">mE_mRNA_A_VirF_1d_head</a> <a href="#">mE_mRNA_A_VirF_4d_head</a> <a href="#">mE_mRNA_A_MateF_20d_head</a> <a href="#">mE_mRNA_A_MateF_4d_head</a> <a href="#">mE_mRNA_A_MateM_20d_head</a> <a href="#">mE_mRNA_A_MateM_4d_acc_gland</a> <a href="#">mE_mRNA_A_MateM_4d_head</a> <a href="#">mE_mRNA_A_MateM_4d_testis</a> <a href="#">mE_mRNA_A_1d_carcass</a> <a href="#">mE_mRNA_A_1d_dig_sys</a> <a href="#">mE_mRNA_A_20d_carcass</a> <a href="#">mE_mRNA_A_20d_dig_sys</a> <a href="#">mE_mRNA_A_4d_carcass</a> <a href="#">mE_mRNA_A_4d_dig_sys</a> <a href="#">mE_mRNA_P8_CNS</a> <a href="#">mE_mRNA_L3_CNS</a> <a href="#">mE_mRNA_L3_Wand_carcass</a> <a href="#">mE_mRNA_L3_Wand_dig_sys</a> <a href="#">mE_mRNA_L3_Wand_fat</a> <a href="#">mE_mRNA_L3_Wand_imag_disc</a> <a href="#">mE_mRNA_L3_Wand_saliv</a> <a href="#">mE_mRNA_A_VirF_20d_head</a> <a href="#">mE_mRNA_A_VirF_4d_ovary</a> <a href="#">mE_mRNA_WPP_fat</a> <a href="#">mE_mRNA_WPP_saliv</a> <a href="#">mE_mRNA_P8_fat</a>	
Number in collection	3,091,281,729	<a href="#">(FlyBase, 2011)</a>
Comment on number in collection	Sum of unique reads of component collections; determined by FlyBase. <a href="#">(FlyBase, 2011)</a>	
Members		

Experimental protocol		
Vector		
Sample preparation	Collection of tissues and preparation of total RNA samples described in modMine entry modENCODE_3207 (link below). <a href="#">(Graveley et al., 2011.4.13)</a>	
Collection preparation	Polyadenylated RNAs were purified from total RNA extracts via oligo(dT) binding, using standard Illumina protocol. The poly(A)+ RNA was fragmented using divalent cations under elevated temperature, following by sequential ligation of RNA linkers to the 5' and 3' ends. Next, reverse transcription was performed using a primer complementary to the 3' linker and PCR was performed using primers complementary to both linkers. ~ 300 bp fragments were isolated from an agarose gel and gel-purified again. <a href="#">(Graveley et al., 2011.4.13)</a>	
Mode of assay	The samples were quantitated using a Nanodrop, and loaded onto a flow cell for cluster generation and sequenced on an Illumina Genome Analyzer II or HiSeq 2000 using paired end protocols (Illumina). <a href="#">(Graveley et al., 2011.4.13)</a>	
Assay platform	Illumina Genome Analyzer II ( <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPI9061">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPI9061</a> ) Illumina HiSeq 2000 <a href="#">(Graveley et al., 2011.4.13)</a>	
Data analysis	Bases were called using the most current version of the Illumina processing pipeline. The sequencing was performed using 76 nt reads on the GAIIx and 100 nt reads on the HiSeq2000. Fastq files were generated using the most current version of the Illumina pipeline software. The HiSeq reads were trimmed from the 3' ends to be 76 nt long, so all aligned reads are 76 nt. Reads were aligned using Bowtie v0.12.0 to a combined index of the genome and both annotated and predicted splice junctions. Paired-end alignments were further parsed using Spliced-Paired-end-Aligner (SPA, written by Michael Duff) to identify the optimal mapping location for mate-pairs. <a href="#">(Graveley et al., 2011.4.13)</a>	
Additional data		
	<b>More information is available under:</b> modENCODE modMine: modENCODE_3207	
Associated files		
Additional sites		
Synonyms and Secondary IDs		
Reported As		
Symbol Synonym	Drosophila transcriptome tissue data <a href="#">(Graveley et al., 2011.4.13)</a> modENCODE_mRNA-Seq_tissues	
Secondary FlyBase IDs		
References ( 2 )		
Personal communication to FlyBase	Graveley et al., 2011.4.13, The D. melanogaster transcriptome: modENCODE RNA-Seq data for dissected tissues The D. melanogaster transcriptome: modENCODE RNA-Seq data for dissected tissues [FBf0213503]	
FlyBase analysis	FlyBase, 2011, Total read counts for modENCODE RNA-Seq data parental collections. Total read counts for modENCODE RNA-Seq data parental collections. [FBf0214761]	

# Dataset report (no parent report)

FB2014\_01, released January 17th, 2014

**FlyBase** Dataset Chromatin\_types\_NKI.Kc167

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

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

General Information			
Name	Chromatin_types_NKI.Kc167	Species	<i>D. melanogaster</i>
Dataset type	genomic sequence feature	FlyBase ID	FBlc0000187
Source and Content			
Consists of	Demarcation of chromatin domains of five major types based on analysis of binding patterns of 53 chromatin proteins, combined with four histone modification profiles. <a href="#">(Filion et al., 2010)</a>		
Created by			
Available from			
Strain			
Stage and tissue			
Cell Line	Kc167		
<input type="checkbox"/> Recent Updates <input checked="" type="checkbox"/> Description and Members			
Description	Whole-genome DamID binding profiles of 53 chromatin proteins in <i>Drosophila</i> Kc167 cells were generated and/or analyzed. In the same array platform, ChIP-on-chip profiles of histone H3, H1, H3K9me2, H3K27me3, H3K4me2, and H3K79me3 were obtained. These were correlated with gene expression, which was measured by RNA-tag profiling. <a href="#">(Filion et al., 2010)</a>		
Parent collections			
Component collection(s)			
Number in collection	8428 <a href="#">(Filion et al., 2010)</a>		
Comment on number in collection	chromatin domains (five types) <a href="#">(Filion et al., 2010)</a>		
Members			
<input checked="" type="checkbox"/> Experimental protocol			
Vector			
Sample preparation	DamID protocol: DNA adenine methyltransferase is fused to the protein of interest, deposits a stable adenine-methylation footprint at interaction site; a DamID construct was created for each chromatin protein assayed and introduced into Kc167 cells; nuclear localization of the Dam-fusion proteins was verified. <a href="#">(Filion et al., 2010)</a>		
Collection preparation	DamID protocol: genomic DNA was isolated from Kc167 cells carrying the DamID construct; DNA regions that were methylated by the Dam-fusion were specifically amplified and labeled with a fluorescent dye. Gene expression: total RNA was isolated from growing Kc167 cells. <a href="#">(Filion et al., 2010)</a>		
Mode of assay	DamID protocol: incorporating a control for non-specific binding, labeled DNA was hybridized to a genomic tiling array, NimbleGen <i>Drosophila</i> 385k array. Histone-ChIP: Using antibodies specific to the histone variants, ChIP maps of histone H3 and four H3 methylation variants were generated on the same array platform. Gene expression: using the Illumina tag-profiling kit with DpnII, poly(A) RNA tag sequencing was carried out on a Illumina Genome Analyzer II. <a href="#">(Filion et al., 2010)</a>		
Assay platform	Illumina Genome Analyzer II ( <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL9061">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL9061</a> ) <a href="#">(Filion et al., 2010)</a> NimbleGen <i>Drosophila</i> 385k array re4 300bp ( <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8471">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8471</a> ) <a href="#">(Filion et al., 2010)</a>		
Data analysis	DamID protocol: arrays were scanned at 5um resolution and raw data were extracted using NimbleScan software; the identity of hybridized material was tracked by the presence of unique oligonucleotide spikes in each sample. Gene expression: tags were mapped by BLAST; only tags mapping to the last GATC of a transcript (FlyBase r5.8) were counted; counts were normalized to the total number of reads. <a href="#">(Filion et al., 2010)</a>		
<input checked="" type="checkbox"/> Additional data			
	More information is available under: <a href="#">GEO: GSE22089</a>		
Associated files			
Additional sites			
<input checked="" type="checkbox"/> Comments			

cell line →

description →

methods {

links to data repositories {



# Dataset report (no parent report)

FB2014\_01, released January 17th, 2014

**FlyBase** Dataset Chromatin\_types\_NKI.Kc167

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

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


General Information			
Name	Chromatin_types_NKI.Kc167	Species	<i>D. melanogaster</i>
Dataset type	genomic sequence feature	FlyBase ID	FBic0000187
Source and Content			
Consists of	Demarcation of chromatin domains of five major types based on analysis of binding patterns of 53 chromatin proteins, combined with four histone modification profiles. <a href="#">(Filion et al., 2010)</a>		
Created by			
Available from			
Strain			
Stage and tissue			
Cell Line	Kc167		
Recent Updates			
Description and Members			
Description	Whole-genome DamID binding profiles of 53 chromatin proteins in <i>Drosophila</i> Kc167 cells were generated and/or analyzed. In the same array platform, ChIP-on-chip profiles of histone H3, H1, H3K9me2, H3K27me3, H3K4me2, and H3K79me3 were obtained. These were correlated with gene expression, which was measured by RNA-tag profiling. <a href="#">(Filion et al., 2010)</a>		
Parent collections			
Component collection(s)			
Number in collection	8428	<a href="#">(Filion et al., 2010)</a>	
Comment on number in collection	chromatin domains (five types)		<a href="#">(Filion et al., 2010)</a>
Members			
Additional data			
	<b>More information is available under:</b>		<a href="#">t al., 2010)</a>
	<b>GEO: GSE22069</b>		<a href="#">t al., 2010)</a>
Associated files			<a href="#">t al., 2010)</a>
Additional sites			
Assay platform	sequencing was carried out on a Illumina Genome Analyzer II. Illumina Genome Analyzer II ( <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL9081">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL9081</a> ) <a href="#">(Filion et al., 2010)</a> NimbleGen <i>Drosophila</i> 385k array re4 300bp ( <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8471">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8471</a> ) <a href="#">(Filion et al., 2010)</a>		
Data analysis	DamID protocol: arrays were scanned at 5um resolution and raw data were extracted using NimbleScan software; the identity of hybridized material was tracked by the presence of unique oligonucleotide spikes in each sample. Gene expression: tags were mapped by BLAST; only tags mapping to the last GATC of a transcript (FlyBase r5.B) were counted; counts were normalized to the total number of reads. <a href="#">(Filion et al., 2010)</a>		
Additional data			
	<b>More information is available under:</b>		
	<b>GEO: GSE22069</b>		
Associated files			
Additional sites			
Comments			

cell line →

description →

links to data repositories {

# Dataset for sequence features – list of members


FB2014\_01, released January 17th, 2014  
Dataset mE1\_TFBS\_HSA

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

General Information			
Name	mE1_TFBS_HSA	Species	<i>D. melanogaster</i>
Dataset type	genomic sequence feature	FlyBase ID	FBic0000258
Source and Content			
Consists of	Genomic sequences identified by integrative analysis of several ChIP data sets.		<a href="#">(The modENCODE Consortium, 2010)</a>
Created by			
Available from			
Strain			
Stage and tissue			
Stage	Tissue/Position (including subcellular localization)	Reference	
<a href="#">embryonic stage</a>		<a href="#">(Negre et al., 2011)</a>	
<b>Comment:</b> TF binding profiles used in this analysis were assayed at early embryo stages.			
Cell Line			
Recent Updates			
Description and Members			
Description	Genomic sequences identified as unique regions of transcription factor (TF) binding using HOT spot analysis (HSA); one or many TFs may bind in a given region. A synthesis of ChIP data sets for 41 different transcription factors.		<a href="#">(The modENCODE Consortium, 2010)</a>
Number in collection	38562		
Comment on number in collection			
Members	<input type="button" value="Download list of the Large dataset metadata members"/>		

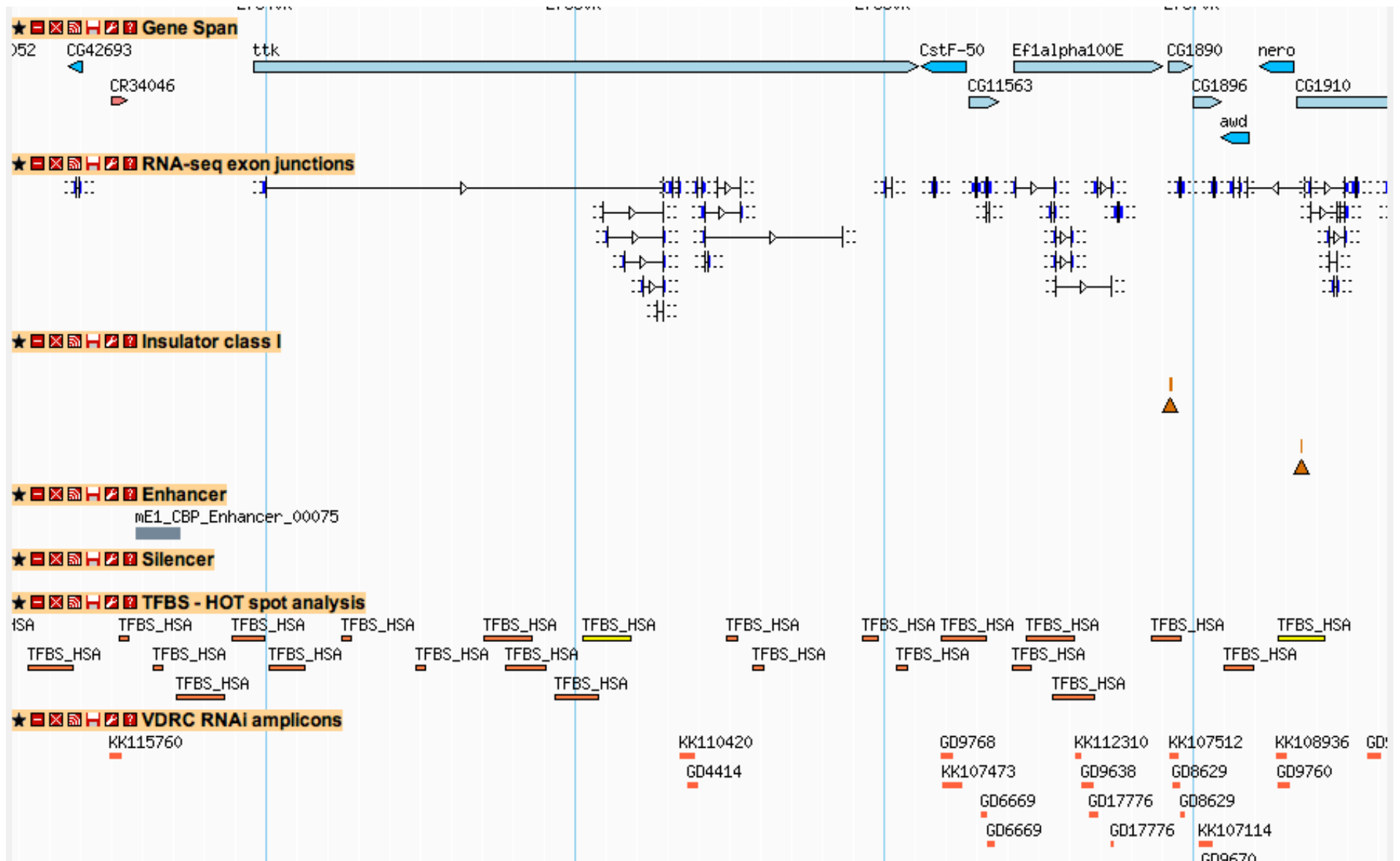
# High-throughput Data: Views, reports, and queries

- GBrowse options (GBrowse 2)
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- **Sequence feature reports**
- FeatureMapper
- Querying RNA-Seq expression data

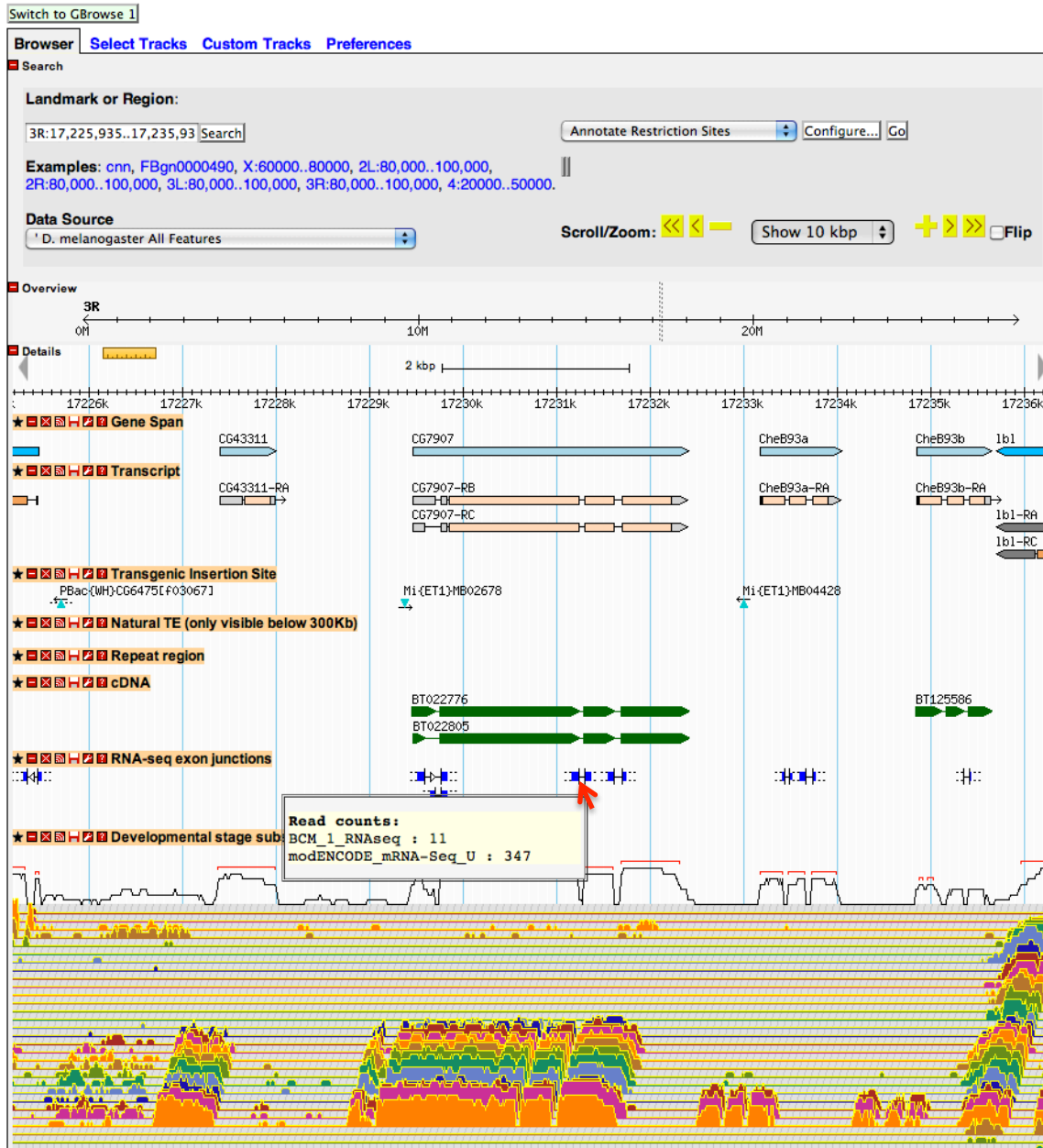
# Sequence Features

What are sequence features?

- Defined sequence range
- Function



# Sequence Features



Sequence Feature Glyph

Hover: popup

Click: navigate to report

Example: RNA-Seq exon junctions

# Sequence Feature reports: RNA-Seq Junctions

FB2014\_01, released January 17th, 2014

**FlyBase** Sequence Feature: exon\_junction

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

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

General Information			
Symbol	Dmel:r5:2R:20269379:20270713:-	Species	<i>D. melanogaster</i>
Feature type	<a href="#">exon_junction</a>	FlyBase ID	FBsf0000131312
Collection	<a href="#">BCM_1_RNAseq_junctions</a> <a href="#">modENCODE_mRNA-Seq_U_junctions</a>	Associated gene(s)	<a href="#">mAChR-A</a>
Genomic Location			
Chromosome (arm)	2R	Sequence location	2R:20,269,379..20,270,713 [-]
Map ( <a href="#">GBrowse</a> ) ?			
<input type="checkbox"/> Recent Updates <input checked="" type="checkbox"/> Sequence Data			
Length			
Comments	GT..AG		
Sequence			
<input checked="" type="checkbox"/> Associated Information <input checked="" type="checkbox"/> Experimental Data <input checked="" type="checkbox"/> RNA-Seq Read Counts			
	<input checked="" type="checkbox"/> Library: <a href="#">BCM_1_RNAseq_junctions</a> <input checked="" type="checkbox"/> Library: <a href="#">modENCODE_mRNA-Seq_U_junctions</a>		
	<input checked="" type="checkbox"/> Progenitors		
<input checked="" type="checkbox"/> Collection Information			
	<input checked="" type="checkbox"/> Collection: <a href="#">BCM_1_RNAseq_junctions</a> <input checked="" type="checkbox"/> Collection: <a href="#">modENCODE_mRNA-Seq_U_junctions</a>		
<input checked="" type="checkbox"/> Comments			
<input checked="" type="checkbox"/> Stocks Listed in FlyBase ( 0 )			
<input checked="" type="checkbox"/> External Crossreferences and Linkouts			
<input checked="" type="checkbox"/> Synonyms and Secondary IDs ( 3 )			
<input checked="" type="checkbox"/> References ( 4 )			

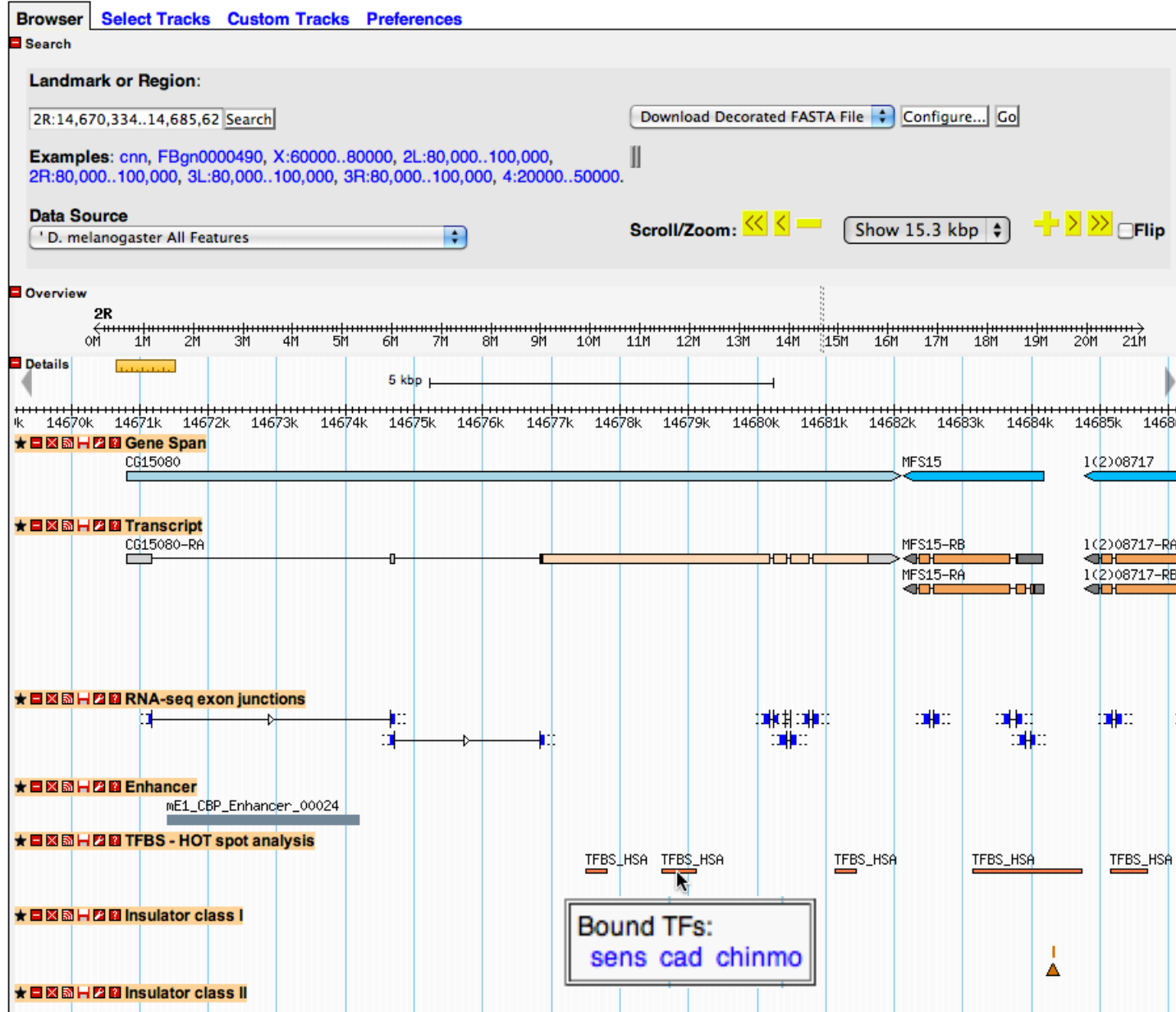


# Sequence Feature reports: RNA-Seq Junctions

embryonic stage 16 <b>Comment:</b> 14-16 hr	7
embryonic stage 17(i) <b>Comment:</b> 16-18 hr	23
embryonic stage 17(ii) <b>Comment:</b> 18-20 hr	26
embryonic stage 17(iii) <b>Comment:</b> 20-22 hr	23
embryonic stage 17(iv) <b>Comment:</b> 22-24 hr	47
first instar larval stage	55
second instar larval stage	18
early third instar larval stage <b>Comment:</b> 12 hr post-molt	1
third instar - uncleared gut stage <b>Comment:</b> indicated as puff stage 1-2	0

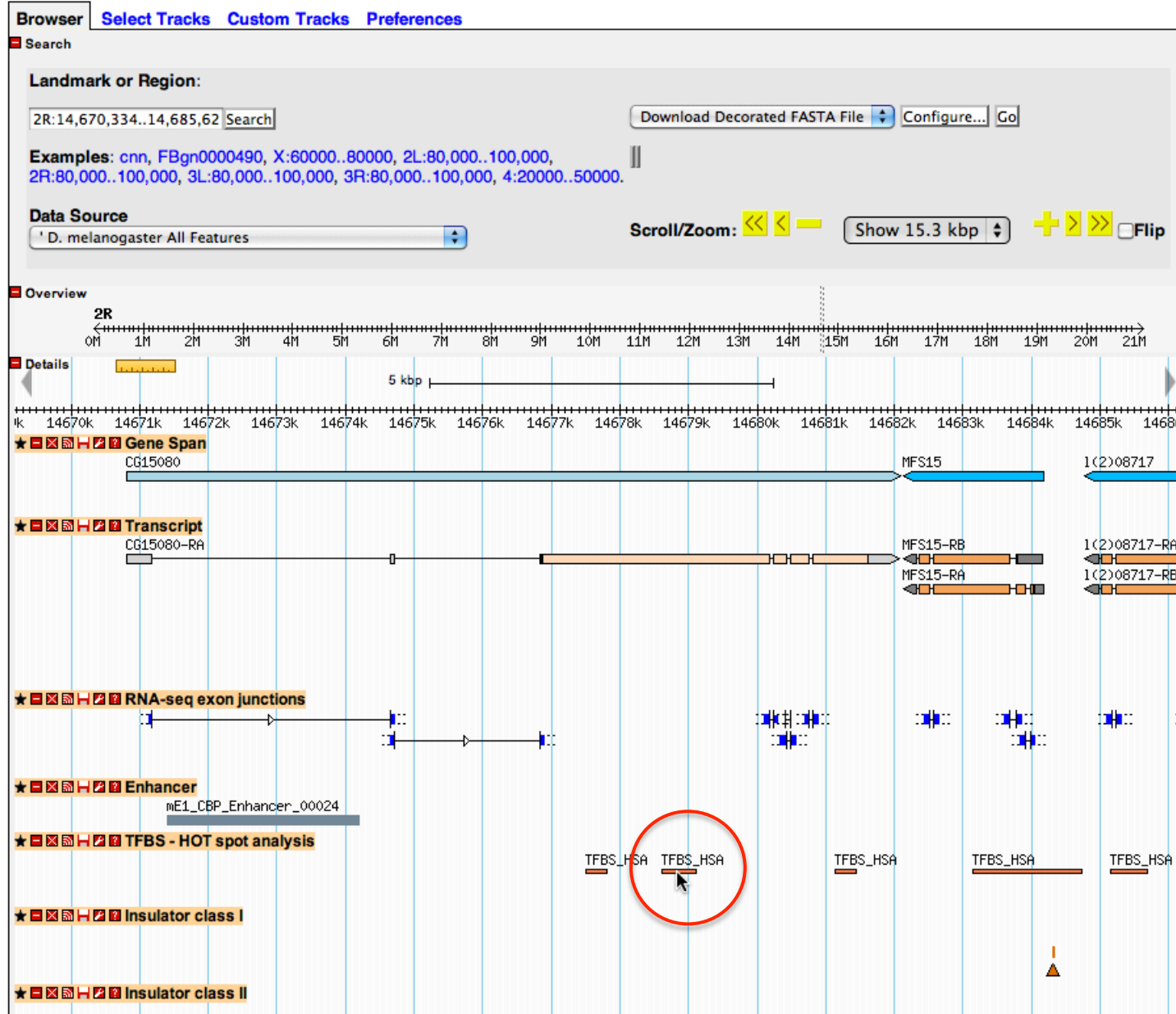
Library: modENCODE_mRNA-Seq_U_junctions	Expression stage(s)	Read Count
embryonic stage 1 -- 4 <b>Comment:</b> 0-2 hr		0
embryonic stage 4 -- 9 <b>Comment:</b> 2-4 hr		0
embryonic stage 9 -- 11 <b>Comment:</b> 4-6 hr		0
embryonic stage 11 -- 12 <b>Comment:</b> 6-8 hr		0
embryonic stage 12 -- 13 <b>Comment:</b> 8-10 hr		0
embryonic stage 13 -- 15 <b>Comment:</b> 10-12 hr		0
embryonic stage 15 -- 16 <b>Comment:</b> 12-14 hr		0
embryonic stage 16 <b>Comment:</b> 14-16 hr		7
embryonic stage 17(i) <b>Comment:</b> 16-18 hr		23
embryonic stage 17(ii) <b>Comment:</b> 18-20 hr		26
embryonic stage 17(iii) <b>Comment:</b> 20-22 hr		23
embryonic stage 17(iv) <b>Comment:</b> 22-24 hr		47
first instar larval stage		55
second instar larval stage		18
early third instar larval stage <b>Comment:</b> 12 hr post-molt		1
third instar - uncleared gut stage <b>Comment:</b> indicated as puff stage 1-2		0
third instar - partially cleared gut stage <b>Comment:</b> indicated as puff stage 3-6		1
third instar - cleared gut stage <b>Comment:</b> indicated as puff stage 7-9		0
prepupal stage P1		2
pupal stage P5 <b>Comment:</b> aged 12 hrs after collection as white prepupae		4
pupal stage P6 <b>Comment:</b> aged 24 hrs after collection as white prepupae		3
pharate adult stage P8 <b>Comment:</b> aged 2 days after collection as white prepupae		6
pharate adult stage P9 -- P10 <b>Comment:</b> aged 3 days after collection as white prepupae		8
pharate adult stage P15 <b>Comment:</b> aged 4 days after collection as white prepupae		13
day 1 of adulthood   female <b>Comment:</b> aged 1 day after eclosion		6
day 1 of adulthood   male <b>Comment:</b> aged 1 day after eclosion		6
day 5 of adulthood   female <b>Comment:</b> aged 5 days after eclosion		2
day 5 of adulthood   male <b>Comment:</b> aged 5 days after eclosion		10
day 30 of adulthood   female <b>Comment:</b> aged 30 days after eclosion		4
day 30 of adulthood   male <b>Comment:</b> aged 30 days after eclosion		12
<b>Total</b>		277

# Sequence Feature reports: TF Binding Sites





# Sequence Feature reports: TF Binding Sites



# Sequence Feature reports: TF Binding Sites

FB2014\_01, released January 17th, 2014

**FlyBase**      Sequence Feature: TF\_binding\_site

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

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

General Information	
Symbol	TFBS_HSA_011820
Species	<i>D. melanogaster</i>
Feature type	TF_binding_site
FlyBase ID	FBs0000358226
Collection	mE1_TFBS_HSA
Associated gene(s)	
Genomic Location	
Chromosome (arm)	2R
Sequence location	2R:14,678,610..14,679,112
Map ( <a href="#">GBrowse</a> )	<p>The map shows a genomic region on chromosome 2R from 14,678,610 to 14,679,112. Key features include:         <ul style="list-style-type: none"> <li>CG15080: A large blue arrow pointing right.</li> <li>eORI_mE_247: A green horizontal bar.</li> <li>TFBS_sens: Two cyan horizontal bars.</li> <li>TFBS_cad: Two black horizontal bars.</li> <li>TFBS_chinmo: A purple horizontal bar.</li> <li>TFBS_HSA: Two red horizontal bars.</li> </ul> </p>
<ul style="list-style-type: none"> <li><input type="checkbox"/> Recent Updates</li> <li><input type="checkbox"/> Sequence Data</li> <li><input type="checkbox"/> Associated Information</li> <li><input type="checkbox"/> Experimental Data</li> <li><input type="checkbox"/> Progenitors</li> <li><input type="checkbox"/> Binding data</li> </ul>	
Bound moiety	<ul style="list-style-type: none"> <li><a href="#">sens protein</a></li> <li><a href="#">cad protein</a></li> <li><a href="#">chinmo protein</a></li> </ul>
Comments on moiety	complexity score = 2.92
Mapped to restriction fragment	
Evidence or Assay	
Collection Information	
Collection: mE1_TFBS_HSAmodENCODE_TFBSs	
Symbol	mE1_TFBS_HSAmodENCODE_TFBSs
Source and Progenitors	
Species of derivation	<i>D. melanogaster</i>
Strain of derivation	
Stage	Tissue/Position (including subcellular localization)      Reference
embryonic stage	<a href="#">(Negre et al., 2011)</a>
Comment: TF binding profiles used in this analysis were assayed at early embryo stages.	
Vector or progenitor construct	
Description	
Genomic sequences identified as unique regions of transcription factor (TF) binding using HOT spot analysis (HSA); one or many TFs may bind in a given region. A synthesis of ChIP data sets for 41 different transcription factors. <span style="float: right;"><a href="#">(The modENCODE Consortium, 2010)</a></span>	
Additional data	

# Sequence Feature reports: TF Binding Sites

FB2014\_01, released January 17th, 2014

**FlyBase** Sequence Feature: TF\_binding\_site

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

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

General Information	
Symbol	TFBS_HSA_011820
Species	<i>D. melanogaster</i>
Feature type	TF_binding_site
FlyBase ID	FBs0000358226
Collection	mE1_TFBS_HSA
Associated gene(s)	
Genomic Location	
Chromosome (arm)	2R
Sequence location	2R:14,678,610..14,679,112
Map ( <a href="#">GBrowse</a> )	<p>The map shows a genomic region on chromosome 2R from 14,678,610 to 14,679,112. Key features include:         <ul style="list-style-type: none"> <li>CG15080 (green arrow pointing right)</li> <li>eORI_mE_247 (cyan bar)</li> <li>TFBS_sens (two blue bars)</li> <li>TFBS_cad (two black bars)</li> <li>TFBS_chinmo (purple bar)</li> <li>TFBS_HSA (two red bars)</li> </ul> </p>
<ul style="list-style-type: none"> <li><input type="checkbox"/> Recent Updates</li> <li><input type="checkbox"/> Sequence Data</li> <li><input type="checkbox"/> Associated Information</li> <li><input type="checkbox"/> Experimental Data</li> <li><input type="checkbox"/> Progenitors</li> <li><input type="checkbox"/> Binding data</li> </ul>	
Bound moiety	<ul style="list-style-type: none"> <li><a href="#">sens protein</a></li> <li><a href="#">cad protein</a></li> <li><a href="#">chinmo protein</a></li> </ul>
Comments on moiety	complexity score = 2.92
Mapped to restriction fragment	
Evidence or Assay	
Collection Information	
Collection: mE1_TFBS_HSAmodENCODE_TFBSs	
Symbol	mE1_TFBS_HSAmodENCODE_TFBSs
Source and Progenitors	
Species of derivation	<i>D. melanogaster</i>
Strain of derivation	
Stage	Tissue/Position (including subcellular localization) Reference
embryonic stage	( <a href="#">Negre et al., 2011</a> )
Comment: TF binding profiles used in this analysis were assayed at early embryo stages.	
Vector or progenitor construct	
Description	
Genomic sequences identified as unique regions of transcription factor (TF) binding using HOT spot analysis (HSA); one or many TFs may bind in a given region. A synthesis of ChIP data sets for 41 different transcription factors. ( <a href="#">The modENCODE Consortium, 2010</a> )	
Additional data	

# Routes to Dataset reports

- GBrowse help links
- Sequence feature reports
- Reference reports
- QuickSearch: Data Class “large dataset metadata”
- QueryBuilder: Large dataset metadata

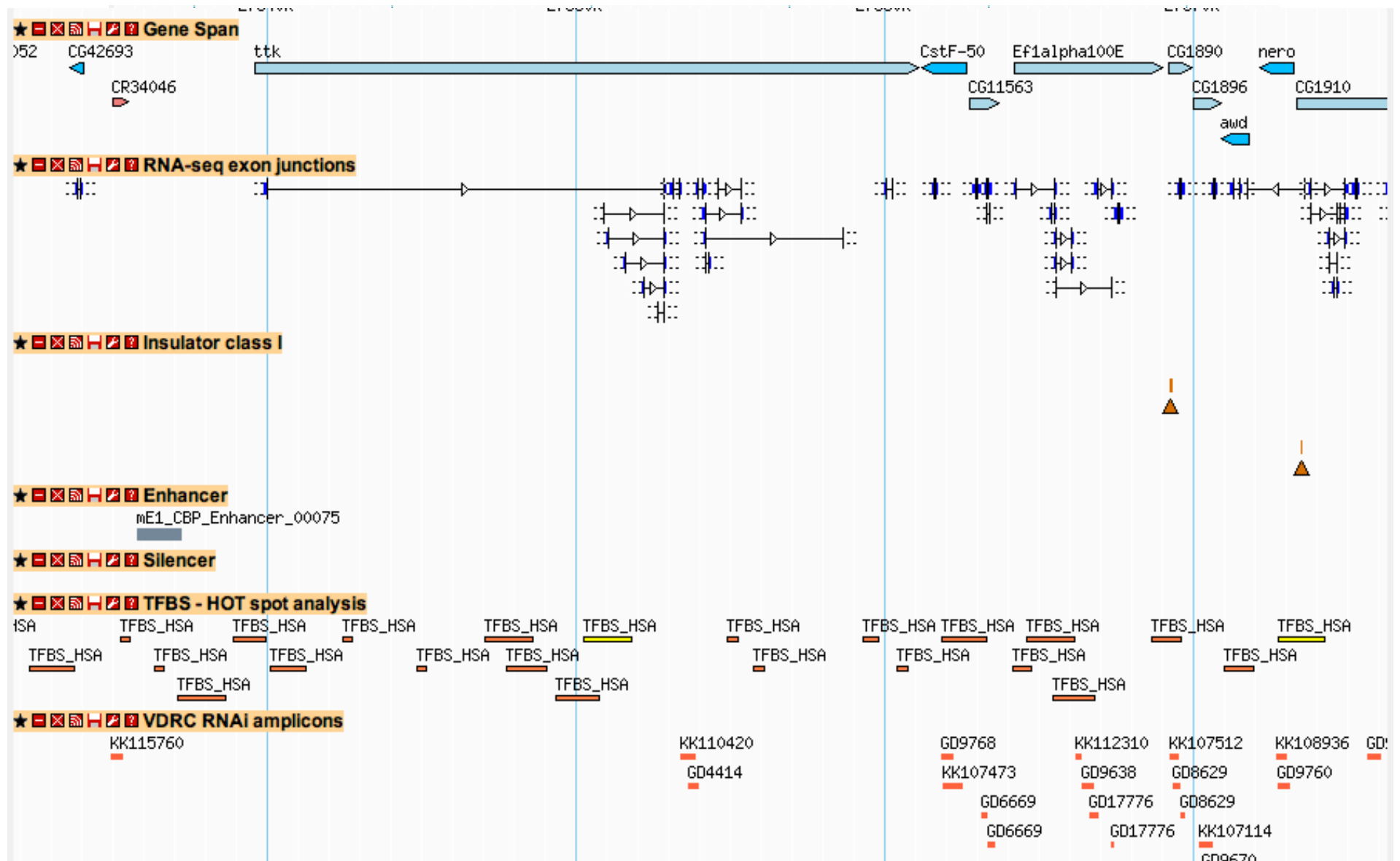
# Routes to Sequence feature reports

- GBrowse links
- Download list from Dataset report
- FeatureMapper
- QuickSearch: Data Class “sequence features”
- QueryBuilder: Sequence Features
- Construct report (for RNAi amplicons)

# High-throughput Data: Views, reports, and queries

- GBrowse options (GBrowse 2)
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- Sequence feature reports
- **FeatureMapper**
- Querying RNA-Seq expression data

# FeatureMapper: find and retrieve features in region of interest



# FeatureMapper: access via Tools menu

FB2014\_01, released January 17th, 2014

## FlyBase

A Database of *Drosophila* Genes & Genomes

Home **Tools** Files Species Documents Resources News Help Archives

 BLAST	 GBrowse	 QueryBuilder	 RNA-Seq Search	 TermLink	 ImageBrowse	 Batch Download
-----------	-------------	------------------	--------------------	--------------	-----------------	--------------------

**Tools** Files Species

- Tools Overview
- Query Tools ▶
- Genomic/Map Tools ▶
  - BLAST
- Retrieve/Convert ▶
  - GBrowse
- Data Submission ▶
  - GBrowse 2 *Beta*
- People ▶
  - CytoSearch
  - Feature Mapper**
  - Aberrations Maps
  - Chromosome Maps
  - Coordinates Converter



# FeatureMapper input

## Mapping Options

### Reference Landmark(s) or Region(s)

Enter ID, Symbol, annotation ID or Sequence Region:

2R:14,668,804..14,684,099  
Sos  
arm

### Set region type to map

Sequence of the Landmark

Include overlapping (not fully enclosed within query region) features

Species: D. melanogaster

### Map Features:

#### Gene Models

- Genes
- mRNA (transcript)
- exon
- five\_prime\_UTR
- three\_prime\_UTR
- tRNA
- miRNA
- snRNA
- snoRNA
- CDS (polypeptide)
- Natural TE

#### Aligned Evidence

- cDNA
- ESTs
- RNA-seq Exon Junctions
- Peptide Atlas peptides

#### Noncoding Features

- Regulatory Regions
- Insulator class I
- Insulator class II
- Protein binding site
- Enhancers
- Silencers
- TFBS - HOT spot analysis
- TFBS - zinc finger domain
- TFBS - homeodomain
- TFBS - helix-loop-helix domain
- TFBS - BTB/POZ domain
- TFBS - other
- Origin of replication
- RNA Editing Sites
- Putative Brain Enhancers

#### Microarray Features

- Affymetrix v1
- Affymetrix v2

#### Mapped Mutations

- Transgene insertion sites
- Point Mutation
- Sequence Variant
- Uncharacterized Change in Seq.
- Aberration Junction
- Complex Substitution
- Indels
- Rescue Fragment

#### RNAi Reagents and Data

- DGRC-1 amplicons
- DGRC-2 oligos
- DRSC RNAi amplicons
- VDRC RNAi amplicons
- TRiP RNAi amplicons
- BKNA RNAi amplicons
- HFA RNAi amplicons
- NIG-Fly RNAi amplicons

Check all

Uncheck all

Group output features by type  GFF lines output

# FeatureMapper output

2R:14,668,804..14,684,099

<b>All features</b>	<a href="#">get GFF file</a>			
Insulators Class I	No features found			
<b>TFBS - HOT spot analysis</b>				<a href="#">to HitList</a>
	2R:14677507..14677811	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_011819</a>
	2R:14678610..14679112	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_011820</a>
	2R:14681137..14681437	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_011821</a>
	2R:14683151..14684729	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_011822</a>
RNA Editing Sites	No features found			

Sos

<b>All features</b>	<a href="#">get GFF file</a>			
Insulators Class I	No features found			
<b>TFBS - HOT spot analysis</b>				<a href="#">to HitList</a>
	2L:13814179..13815271	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_004500</a>
	2L:13816450..13817860	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_004501</a>
	2L:13818046..13818414	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_004502</a>
	2L:13819475..13820855	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_004503</a>
RNA Editing Sites	No features found			

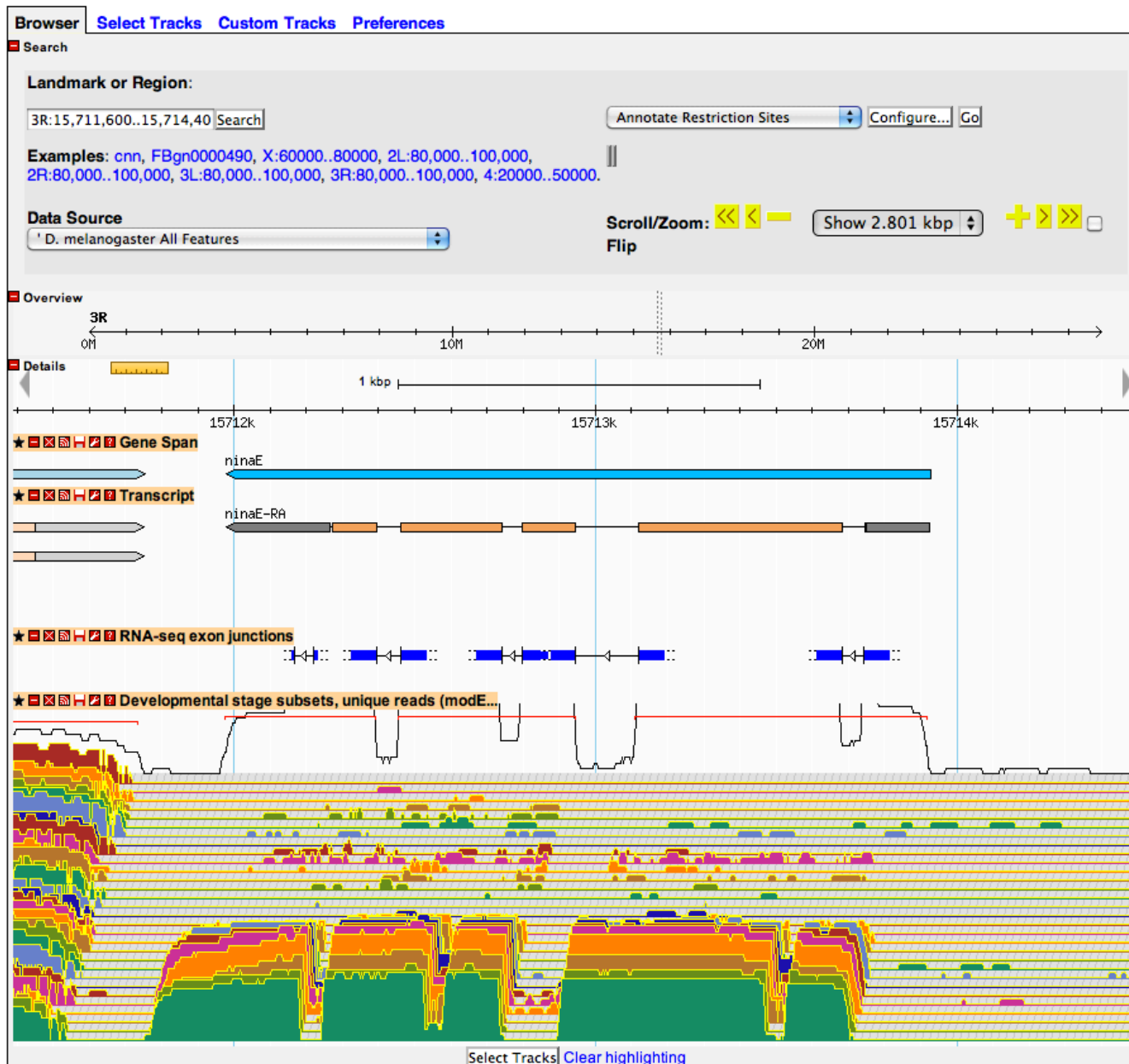
arm

<b>All features</b>	<a href="#">get GFF file</a>			
<b>Insulators Class I</b>				<a href="#">to HitList</a>
	X:1787761..1787771	->	insulator:Insulator_Class_I.mE01	<a href="#">Insulator_I_3962</a>
	X:1794581..1794591	->	insulator:Insulator_Class_I.mE01	<a href="#">Insulator_I_3963</a>
<b>TFBS - HOT spot analysis</b>				<a href="#">to HitList</a>
	X:1786392..1787874	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_031870</a>
	X:1787337..1788789	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_031871</a>
	X:1788101..1788111	->	TF_binding_site:mE1_TFBS_HSA	<a href="#">TFBS_HSA_031872</a>

# High-throughput Data: Views, reports, and queries

- GBrowse options (GBrowse 2)
- Dataset reports (large dataset metadata)
  - Links to GEO and modMine
- Sequence feature reports
- FeatureMapper
- Querying RNA-Seq expression data

# RNA-Seq coverage data in GBrowse



# Querying RNA-Seq expression data

GTGGCAATCGTAAGATAGCCAAATATTATTATTGTTTCAGATACTCAG  
AGGCGCCAAATGCGCAGATGTCGAGAGTGTGTCGAGTGAATCAGTGAJATT  
ATTTCCGCGCAAAAGCGACTTCTTGGGTTGAAAGGAGGAAAGGAAAGGAA  
AATAATAAAAAAACAACAGTGCAGACAGCCGGGGCATCTTCATAGA

FB2014\_01, released January 17th, 2014

## FlyBase

A Database of *Drosophila* Genes & Genomes

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

[BLAST](#) [GBrowse](#) [QueryBuilder](#) [RNA-Seq Search](#) [TermLink](#) [ImageBrowse](#) [Batch Download](#)

**Fast-Track Your Paper**

[FlyBase Forum](#)

[Find a Fly Person](#)

**News**

[New in Release FB2014\\_01](#) | 17 Jan 14  
[2014 Release Schedule](#) | 15 Jan 14  
[FlyBase 102](#) | 1 Dec 13  
[Stock Center User Survey](#) | 18 Jun 13  
[Fly Mating Scheme Design](#) | 13 Mar 13  
[GenBank Release](#) | 25 Jan 13  
[NAR article on bibliography](#) | 5 Nov 12  
[White Paper 2012](#) | 8 Jun 12

**Upcoming Meetings**

[55th ADRC](#) | 26 Mar 14  
[Reg Dros Meeting 2014](#) | 27 May 14  
[19th Crele Dros Mol/Dev Bio](#) | 22 Jun 14  
[7th Intl Symp Mol Insect Sci](#) | 13 Jul 14  
[XV Euro Dros Neurobiology](#) | 5 Oct 14  
[Dros in Exp Genetics & Bio](#) | 6 Oct 14  
[2nd ESF-EMBO Minibrains](#) | 8 Nov 14

**Courses**

[Dros Species Workshop XII](#) | 25 Mar 14  
[W-T Dros Genetics/omics](#) | 27 Jul 14

**QuickSearch**

[Simple](#) [Expression](#) [Phenotype](#) [GO](#) [References](#) [Data Class](#)

[Go to RNA-Seq Search](#)

or

Search literature-curated expression patterns:

Species:  include non-Dmel species

refine search by adding qualifier terms (such as 'ventral') e.g., third instar larval stage

Stage:  e.g., neuron

Tissue:  e.g., growth cone

Cell Loc.:

Note: Wild cards (\*) can be added to your search term

**Commentary** [See all commentaries](#)

**FlyBase RNA-Seq RPKM data bulk download**

Tissue/Stage	Count
digestive system, 1-day adult	2
digestive system, 4-day adult	4
digestive system, 20-day adult	3
fat body, larvae L3 wandering	4
fat body, white prepupae	40
fat body, pupae PB	25
carcass, larvae L3 wandering	15
carcass, 1-day adult	15
carcass, 4-day adult	16
carcass, 20-day adult	11
ovary, virgin 4-day female	66
ovary, mated 4-day female	54
testis, mated 4-day male	65

May 6, 2013. FlyBase is extending its initial gene-level analyses of RNA-seq throughput data from modENCODE and others. The algorithm for RPKM (reads per kilobase per million mapped reads) has been refined, additional datasets have been analyzed, and these data are now available for bulk download... [\(More\)](#)

FlyBase is supported by a grant from the National Human Genome Research Institute at the U.S. National Institutes of Health #P41 HG000739. Support is also provided by the British Medical Research Council, the Indiana Genomics Initiative, and the National Science Foundation through XSEDE resources provided by Indiana University. [Copyright Statement](#)

version FB2014\_01, released January 17, 2014

[Contact FlyBase](#) [Cite FlyBase](#) [Site Map](#)

[Twitter](#) [YouTube](#) [@](#)

# Querying RNA-Seq expression data

The forms below can be used to query FlyBase records using the modENCODE hit

**search using several modENCODE expression datasets in conjunction**

Join selections in the following forms for the search:  stage  tissue  treatr  cell lir

**modENCODE expression by tissue data**

"Expression off" means a peak expression level<sup>1</sup> not more than  expression

Guide to modENCODE expression RPKM level bins*	
No/Extremely low	0 - 0
Very low	1 - 3
Low	4 - 10
Moderate	11 - 25
Moderately high	26 - 50
High	51 - 100
Very high	101 - 1000
Extremely high	>1000

\*Gelbart and Emmert, 2013

Graveley et al., 2010.03.15

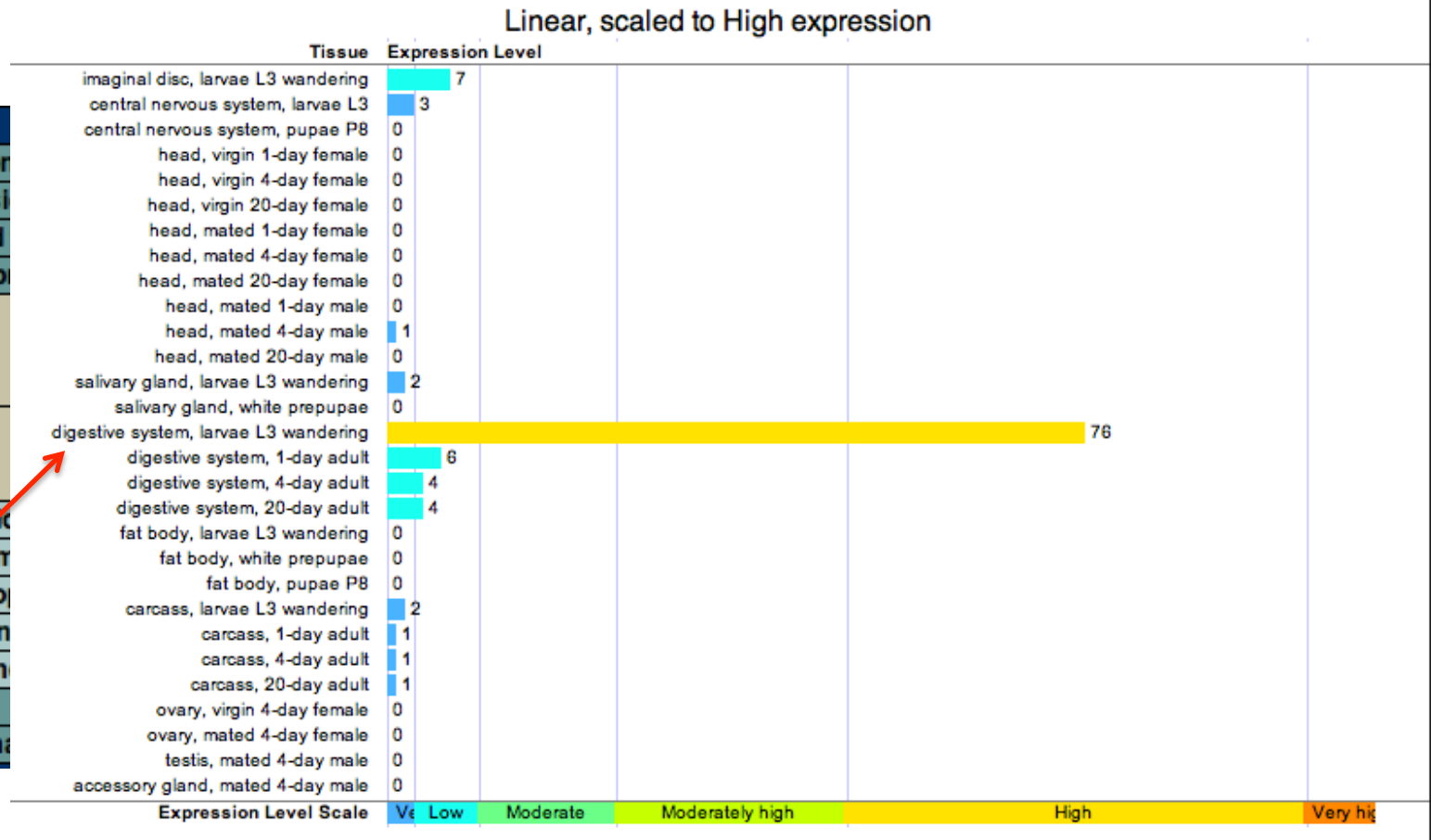
[Documentation](#)

89 matches    Convert to     Results Analysis/Refinement    HitList Conversion Tools

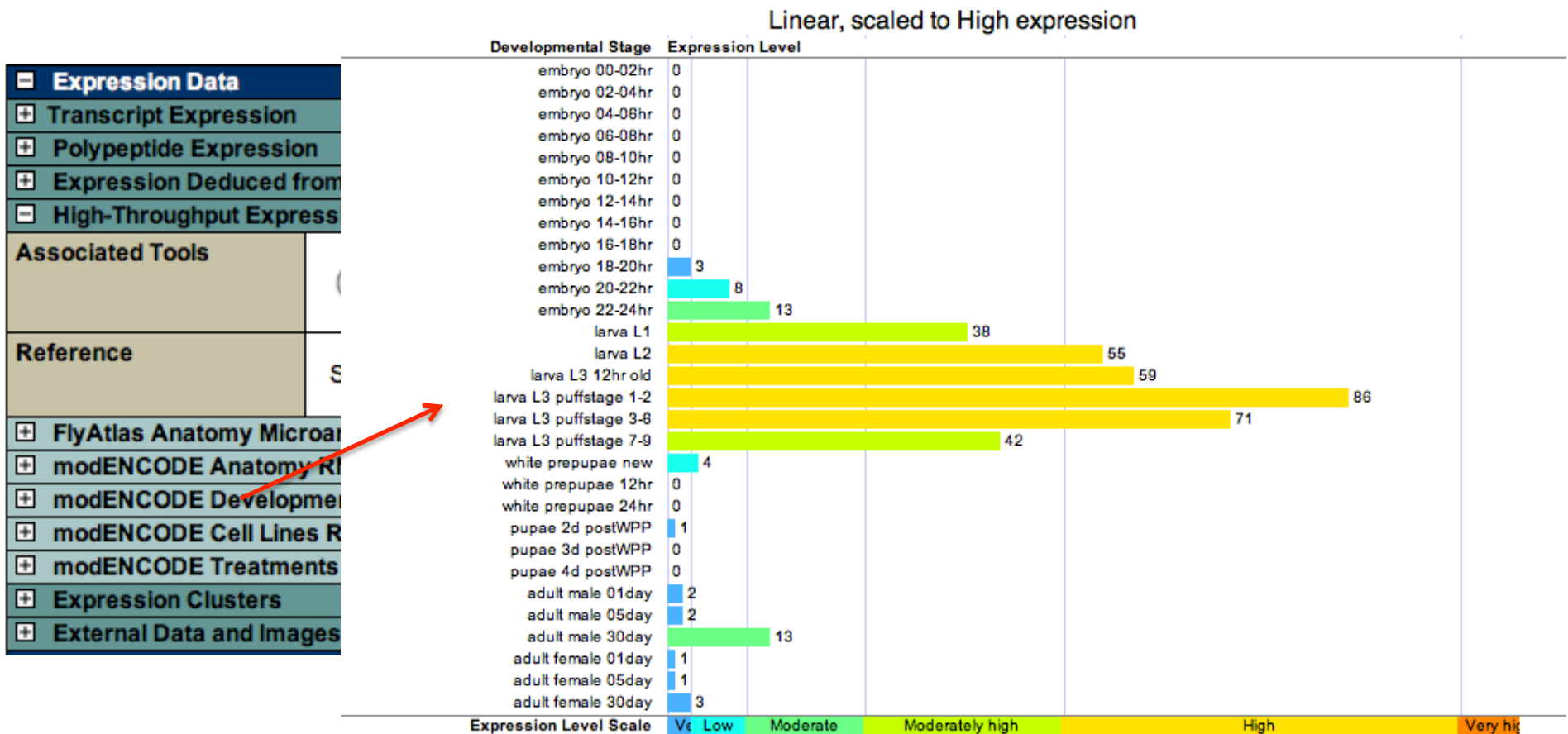
<input checked="" type="checkbox"/>	#	Symbol ↑	Name ↑	Annotation ID ↑	Cytology ↑	Alleles # ↑	Stocks # ↑	Clones # ↑
<input checked="" type="checkbox"/>	1	Cda9	Chitin deacetylase-like 9	CG15918	53F13-53F13	1	1	5
<input checked="" type="checkbox"/>	2	CG3264	-	CG3264	58D1-58D1	2	2	15
<input checked="" type="checkbox"/>	3	CG3292	-	CG3292	58D1-58D1	5	1	7
<input checked="" type="checkbox"/>	4	CG3734	-	CG3734	92A3-92A3	4	3	96
<input checked="" type="checkbox"/>	5	CG3814	-	CG3814	49D6-49D6	6	5	23
<input checked="" type="checkbox"/>	6	CG4363	-	CG4363	58A2-58A2	3	1	1
<input checked="" type="checkbox"/>	7	CG4377	-	CG4377	58A2-58A2	1	2	2
<input checked="" type="checkbox"/>	8	CG4562	-	CG4562	92B4-92B4	7	6	19
<input checked="" type="checkbox"/>	9	CG5246	-	CG5246	89F1-89F1	2	2	13
<input checked="" type="checkbox"/>	10	CG5550	-	CG5550	53D10-53D10	3	3	18
<input checked="" type="checkbox"/>	11	CG5883	-	CG5883	68E3-68E3	2	1	4
<input checked="" type="checkbox"/>	12	CG6048	-	CG6048	5C10-5C10	5	3	10
<input checked="" type="checkbox"/>	13	CG6839	-	CG6839	75E5-75E6	5	3	9
<input checked="" type="checkbox"/>	14	CG6996	-	CG6996	77A2-77A2	2	2	3
<input checked="" type="checkbox"/>	15	CG7542	-	CG7542	74D4-74D4	4	1	3
<input checked="" type="checkbox"/>	16	CG7968	-	CG7968	34D8-34D8	2	1	2
<input checked="" type="checkbox"/>	17	CG7992	-	CG7992	18B4-18B5	4	4	4
<input checked="" type="checkbox"/>	18	CG8907	-	CG8907	89E8-89E8	6	5	46
<input checked="" type="checkbox"/>	19	CG9084	-	CG9084	47D7-47D7	3	2	121
<input checked="" type="checkbox"/>	20	CG9380	-	CG9380	60F5-60F5	3	3	140
<input checked="" type="checkbox"/>	21	CG9903	-	CG9903	14E1-14E1	2	2	7
<input type="checkbox"/>	22	CG10086	-	CG10086	84E1-84E1	5	4	15
<input type="checkbox"/>	23	CG10116	-	CG10116	70B1-70B1	3	3	8
<input type="checkbox"/>	24	CG10140	-	CG10140	70B1-70B1	2	2	5
<input checked="" type="checkbox"/>	25	CG10405	-	CG10405	89B17-89B17	4	4	12
<input checked="" type="checkbox"/>	26	CG10659	-	CG10659	38B1-38B1	3	2	3
<input checked="" type="checkbox"/>	27	CG11340	-	CG11340	100C1-100C1	6	6	2
<input checked="" type="checkbox"/>	28	CG11671	-	CG11671	84F8-84F8	4	2	4
<input checked="" type="checkbox"/>	29	CG13492	-	CG13492	58A2-58A2	5	5	5
<input checked="" type="checkbox"/>	30	CG13806	-	CG13806	62D4-62D4	4	2	6
<input checked="" type="checkbox"/>	31	CG14205	-	CG14205	18D3-18D3	2	1	8
<input checked="" type="checkbox"/>	32	CG14820	-	CG14820	65D3-65D3	2	2	1
<input checked="" type="checkbox"/>	33	CG14957	-	CG14957	63B10-63B10	2	2	7
<input checked="" type="checkbox"/>	34	CG15152	-	CG15152	36E5-36E5	2	2	8
<input checked="" type="checkbox"/>	35	CG15155	-	CG15155	36E6-36E6	2	2	7
<input checked="" type="checkbox"/>	36	CG15170	-	CG15170	37B8-37B8	5	6	1

# RNA-Seq coverage data: RPKM histograms

- Expression Data
- Transcript Expression
- Polypeptide Expression
- Expression Deduced
- High-Throughput Expression
- Associated Tools
- Reference
- FlyAtlas Anatomy Microarray
- modENCODE Anatomy Microarray
- modENCODE Developmental Time Course
- modENCODE Cell Lines
- modENCODE Treatments
- Expression Clusters
- External Data and Images



# RNA-Seq coverage data: RPKM histograms





# Release 6 is Coming!



- FlyBase is working with NCBI, BDGP and other major data producers to migrate all genome annotation data to Release 6.
- We have begun the migration process and expect it to be completed by the end of the summer of 2014.
- We will coordinate public migration to Release 6 with NCBI and other major genome archives starting in the summer of 2014.
- After Release 5.57 (May 2014), the next gene model annotation set will be on the Release 6 assembly.

# References

## Main modENCODE Publications:

Negre, et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 6(1): e1000814. ([FBrf0209760](#))

modENCODE Consortium, et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330(6012): 1787--1797. ([FBrf0212741](#))

Graveley, et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature 471(7339): 473--479. ([FBrf0213330](#))

Negre, et al. (2011). A cis-regulatory map of the Drosophila genome. Nature 471(7339): 527--531. ([FBrf0213303](#))

Kharchenko, et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471(7339): 480--485. ([FBrf0213294](#))

Berezikov, et al. (2011). Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome Res. 21(2): 203-215. ([FBrf0213080](#))

Chung, et al. (2011). Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. Genome Res. 21(2): 286--300. ([FBrf0213087](#))

Cherbas, et al. (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome Res. 21(2): 301--314. ([FBrf0213077](#))

Eaton, et al. (2011). Chromatin signatures of the Drosophila replication program. Genome Res. 21(2): 164--174. ([FBrf0213065](#))

Riddle, et al. (2011). Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res. 21(2): 147--163. ([FBrf0213099](#))

Brooks, et al. (2011). Conservation of an RNA regulatory map between Drosophila and mammals. Genome Res. 21(2): 193--202. ([FBrf0213079](#))

Smibert, et al. (2012). Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila. Cell Rep. 1(3): 277--289. ([FBrf0218523](#))

## modENCODE Overviews:

Celniker, et al. (2009). Unlocking the secrets of the genome. Nature 459(7249): 927-930. ([FBrf0215491](#))

Contrino, et al. (2012). modMine: flexible access to modENCODE data. Nucleic Acids Res. 40(1): D1082--D1088. ([FBrf0217069](#))

Washington, et al. (2011). The modENCODE Data Coordination Center: lessons in harvesting comprehensive experimental details. Database (Oxford) 2011(): bar023. ([FBrf0214707](#))