



Large-scale queries of FlyBase

Steven Marygold

FlyBase, Dept of PDN, University of Cambridge

sjm41@cam.ac.uk

Cambridge Fly Club Symposium - 28th January 2019

Acknowledgements

FB-Harvard (USA):

Norbert Perrimon (PI)
Susan Russo Gelbart
Julie Agapite
Kris Broll
Lynn Crosby
Gil dos Santos
Sian Gramates
Kathleen Falls
Victoria Jenkins
Beverley Matthews
Carol Sutherland
Christopher Tabone
Pinglei Zhou
Mark Zytovicz

FB-Cambridge (UK):

Nick Brown (PI)
Giulia Antonazzo
Helen Attrill
Phani Garapati
Alex Holmes
Aoife Larkin
Steven Marygold
Gillian Millburn
Clare Pilgrim
Vitor Trovisco
Pepe Urbano

FB-Indiana (USA):

Thom Kaufman (PI)
Brian Calvi (PI)
Josh Goodman
Victor Strelets
Jim Thurmond

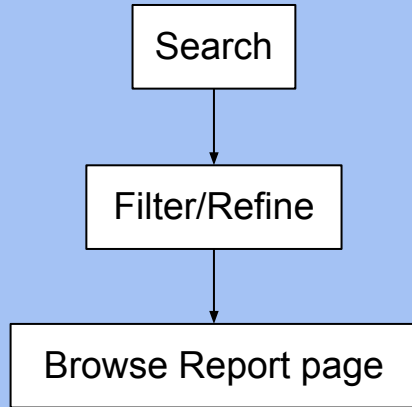
FB-New Mexico (USA):

Richard Cripps (PI)
Maggie Werner-Washburne (PI)
Phillip Baker

Funding:

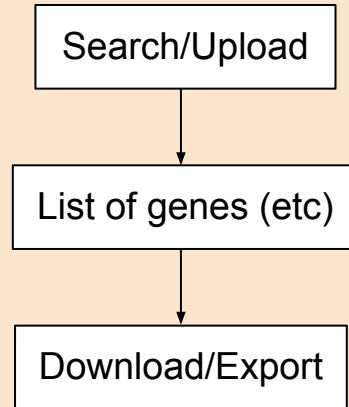
NHGRI at the U.S. National Institutes of Health (U41HG000739)
UK Medical Research Council (MR/N030117/1)
Donations from FlyBase users

FLYBASE WEBSITE



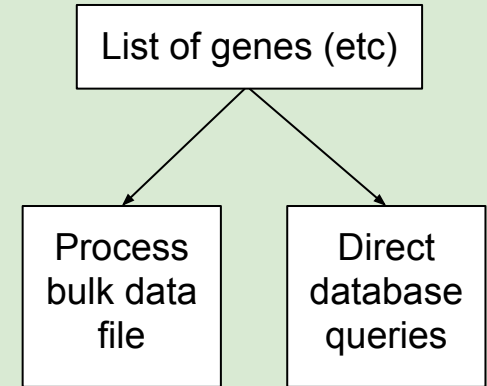
- Single gene queries
- Regular users

FLYBASE WEBSITE



- Multiple gene queries
- Regular users

FTP SITE / SQL / API



- Multiple gene queries
- Power users

Aims

- Raise awareness of methods to obtain data FlyBase in bulk (batch mode)
- Overview of how to use these methods
- Encourage you to try them and provide feedback

Outline

1. Obtaining data on single genes (web)
2. Obtaining data on multiple genes (web)
3. Bulk data files (FTP site)
4. Database downloads, queries & API access
5. Additional information/help
6. Questions / feedback / use cases

Part 1: Obtaining data on single genes (web)

- Jump to Gene/Search FlyBase box
- QuickSearch
- Hitlists (List view)
- Reports

Jump to Gene/Search FlyBase



J2G ▾ Jump to Gene Go



J2G ▾ Cdk1 Go



Gene: Dmel\Cdk1

Search ▾ Search FlyBase Go



Search ▾ Cdk1 Go



HitList

QuickSearch

QuickSearch

[Human Disease](#)

[GAL4 etc](#)

[Expression](#)

[Phenotype](#)

[References](#)

[Search FlyBase](#)

[Orthologs](#)

[Protein Domains](#)

[Gene Groups](#)

[GO](#)

[Data Class](#)

ID, Symbol, Name ▾

ID, Symbol, Name

Everything

Search

Note: [Wild cards](#) (*) can be added to your search term

Hitlists (List view)

145 selected

Items
1-50 of 145

Filter by species

- D. melanogaster* (141)
- H. sapiens* (transgenes in flies) (4)
- other *Drosophila* species (11)
- Other species (43)

Filter by data class

- Aberration (12)
- Allele (100)
- Clone (81)
- Dataset (163)
- Experimental Tool (11)
- Gene (45)
- Gene Group (3)
- Gene Ontology (4)
- Human Disease Model (1)
- Insertion (30)
- Natural Transposon (1)
- Physical Interaction (25)
- Polypeptide (1)
- Reference (413)
- Sequence Feature (18)
- Stock (17)
- Transcript (1)
- Transgenic Construct (54)

[show all](#)

Cdk1 Cyclin-dependent kinase 1 (CG5363, FBgn0004106) *D. melanogaster*

Feature type: protein coding gene
Sequence Location: 2L:10,384,739..10,386,262 [-]

45 Alleles 24 Stocks 1 Transcript 1 Polypeptide 387 References

Gene model status: Current
Cytogenetic Map: 31D11-31D11

Gene Snapshot >

Cdk1^{216A} (FBaI0001571) *D. melanogaster*

Allele class: hypomorphic allele - genetic evidence
Mutagen: ethyl methanesulfonate **Known lesion?** yes

No associated insertions or constructs
5 Phenotype (Class) statements, 2 Phenotype (Anatomy) statements

0 Stocks 3 References

Cdk1^{D57} (FBaI0001572) *D. melanogaster*

Allele class: amorphic allele - genetic evidence
Mutagen: ethyl methanesulfonate **Known lesion?** yes

No associated insertions or constructs
5 Phenotype (Class) statements, 2 Phenotype (Anatomy) statements

0 Stocks 4 References

Cdk1^{216P} (FBaI0030730) *D. melanogaster*


Allele class: hypomorphic allele - genetic evidence
Mutagen: PM hybrid dysgenesis, P-element activity **Known lesion?** yes

Associated insertion: P{}Cdk1^{216P}

6 Phenotype (Class) statements, 3 Phenotype (Anatomy) statements


0 Stocks 5 References

Report Page

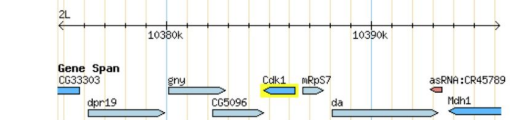
 Gene: Dmel\Cdk1

[Home](#) [Tools](#) [Downloads](#) [Links](#) [Community](#) [Species](#) [About](#) [Help](#) [Archives](#) Search FlyBase

General Information

Symbol	Dmel\Cdk1	Species	<i>D. melanogaster</i>
Name	Cyclin-dependent kinase 1	Annotation Symbol	CG5363
Feature Type	protein_coding_gene	FlyBase ID	FBgn0004106
Gene Model Status	Current	Stock Availability	24 publicly available
Enzyme Name (EC)	Cyclin-dependent kinase (2.7.11.22)		
Gene Snapshot	Cyclin-dependent kinase (Cdk1) is a catalytic protein kinase subunit that can only become active after association with either <i>CycA</i> , <i>CycB</i> or <i>CycB3</i> . The protein kinase activities of these complexes (<i>CycA</i> -Cdk1, <i>CycB</i> -Cdk1, <i>CycB3</i> -Cdk1) control important aspects of progression through the cell cycle. Functionally, the different Cdk1 complexes are partially redundant. They phosphorylate hundreds of target proteins and are most important for progression into and through mitotic and meiotic M phases. [Date last reviewed: 2016-06-23]		
Other Summaries	<input type="button" value="Auto summary"/> <input type="button" value="Gene Group"/> <input type="button" value="UniProtKB"/> <input type="button" value="Interactive Fly"/>		
Also Known As	cdc2, Dmcdc2		
Key Links	 <input type="button" value="NCBI Gene"/> <input type="button" value="Ensembl"/> <input type="button" value="UniProtKB"/>		

Genomic Location

Cytogenetic map	31D11-31D11	Sequence location	2L:10,384,739..10,386,262 [-]
Recombination map	2-41		
Sequence	<input type="text" value="Gene region"/> <input type="button" value="Get Decorated FASTA"/>		
	<input type="text" value="Get Sequence"/>		
Genomic Maps	<input type="button" value="GBrowse"/> <input type="button" value="JBrowse"/>		
Help me choose			
Other Genome Views	The following external sites may use different assemblies or annotations than FlyBase. <input type="button" value="NCBI Genome Data Viewer"/> <input type="button" value="UCSC Genome Browser"/> <input type="button" value="Ensembl Genome Browser"/> <input type="button" value="PopFly Genome Browser"/>		

Report Sections

- General Information
- Genomic Location
- GO Summary Ribbons
- Families, Domains and Molecular Function
- Summaries
- Gene Model and Products**
- Gene Ontology
- Expression Data
- Alleles, Insertions, Constructs, and Phenotypes
- Orthologs
- Human Disease Model Data
- Functional Complementation
- Interactions
- Pathways
- Genomic Location and Mapping
- Stocks and Reagents
- Other Information
- Crossreferences
- Synonyms and Secondary IDs
- Datasets
- References

Part 2: Obtaining data on multiple genes (web)


- Gene lists in FlyBase; Upload/Convert IDs
- Hitlists (Table view); HitList tools
- Export to Batch Download
- Export to FeatureMapper
- Sequence Downloader

Gene lists in FlyBase


Three main ways to generate/obtain a list:

1. From a search (and subsequent refinement) in FlyBase
2. Via a link within a FlyBase report page
3. By uploading your own list

Hitlists via FlyBase report pages

 **FlyBase** Gene Group : INTEGRINS

Home Tools Downloads Links Community Species About Help Archives Search Search FlyBase Go

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

General Information

Name	INTEGRINS	Species	<i>D. melanogaster</i>
Symbol	ITG	FlyBase ID	FBgg0000058
Date last reviewed	2013-11-11	Number of members	7

Description

Description Integrins are heterodimeric transmembrane receptors composed of an α and β subunit that mediate cell-cell and cell-extracellular matrix adhesion. As well as maintaining tissue integrity, they are involved in morphogenesis and development. (Adapted from FBr0167428).

Notes on Group

Source Material The INTEGRINS Gene Group has been compiled by FlyBase curators using the following publication(s): [Brower, 2003](#).

Key Gene Ontology (GO) terms

Molecular Function	protein heterodimerization activity
Biological Process	cell adhesion mediated by integrin
Cellular Component	integrin complex integral component of membrane

Related Gene Groups

Members (7)


For all members: [View Orthologs](#) [Export to HitList](#) [Export to Batch Download](#)

Gene Symbol	Gene Name	Also Known As	Source Material for Membership
if	inflated	α PS2, α ps2, PS2, PS2 α , PS 2	(Brower, 2003)
ItgaPS4	Integrin alphaPS4 subunit	α PS4, alphaPS4, ItgaPS4, α -ps4	(Brower, 2003)
ItgaPS5	Integrin alphaPS5 subunit	α PS5, alphaPS5, α -PS5, ItgaPS5	(Brower, 2003)
Itgbn	Integrin betanu subunit	β v, betaInt-nu, β Int-v, β v, β v	(Brower, 2003)
mew	multiple edematous wings	α PS1, α ps1, PS1, PS1 α , α 1Int	(Brower, 2003)
mys	mysospheroid	β PS, β ps, I(1)mys, olfC, β PS-integrin	(Brower, 2003)
scb	scab	α PS3, Vol, scab, I(2)51Ea, α ps3	(Brower, 2003)

Report Sections

- [General Information](#)
- [Description](#)
- [Members](#)
- [External Data](#)
- [Synonyms and Secondary IDs](#)
- [References](#)

Upload/Convert IDs



FB2018_06, relea

Home **Tools** Downloads Links Community Spec

- Tools Overview & Help
- Query by symbols/IDs**
- Search/Browse Portals
- Genomics Tools
- Submit Data
- Researcher Directory

- Batch Download
- Sequence Downloader
- Upload/Convert IDs**



Upload/Convert IDs

[Help](#)

Validate Only (Update to Current IDs)

Validate and Convert into: Genes

Enter IDs or Symbols:

You may enter (or upload) FlyBase IDs, symbols, annotation symbols (CG#), clone names, PubMed IDs, or GenBank/Uniprot/Swiss-Prot/TrEMBL accessions. Please use spaces or returns to separate the identifiers (no commas or other

or Upload File of IDs:

Browse

Submit Query

Reset

Upload/Convert IDs

[Help](#)

Validate Only (Update to Current IDs)

Validate and Convert into: Genes

Enter IDs or Symbols:

FBgn0001250
FBgn0004456
mys
FBgn0010395
CG16827
Bgn0034880

or Upload File of IDs:

Browse

Submit Query

Reset

Submitted IDs: 7 | Validated/Updated IDs: 6 | Unknown IDs: 1 | Unique converted IDs: 6
Export converted IDs to: [HitList](#) [BatchDownload](#) Save as: [file, uniq IDs](#) [file, conversion table](#)

Conversion report

Submitted ID	Current ID	Converted ID	Related record
Bgn0034880 - unknown ID			
FBgn0011707	FBgn0286785	FBgn0286785	scb
FBgn0004456	FBgn0004456	FBgn0004456	mew
CG16827	FBgn0034005	FBgn0034005	ltgaPS4
FBgn0010395	FBgn0010395	FBgn0010395	ltgbn
mys	FBgn0004657	FBgn0004657	mys
FBgn0001250	FBgn0001250	FBgn0001250	if

Hitlists - Table view

[Convert](#) [Export](#) [Analyze](#)

View As
[List](#) [Table](#)

Filter by species [clear](#)

- D. melanogaster* (7)
- H. sapiens* (transgenes in flies) (0)
- other *Drosophila* species (0)
- Other species (0)

Filter by data class [clear](#)

- Gene (7)

[show all](#)

7 selected [New Hitlist](#)

[← Previous](#) [1](#) [Next →](#)

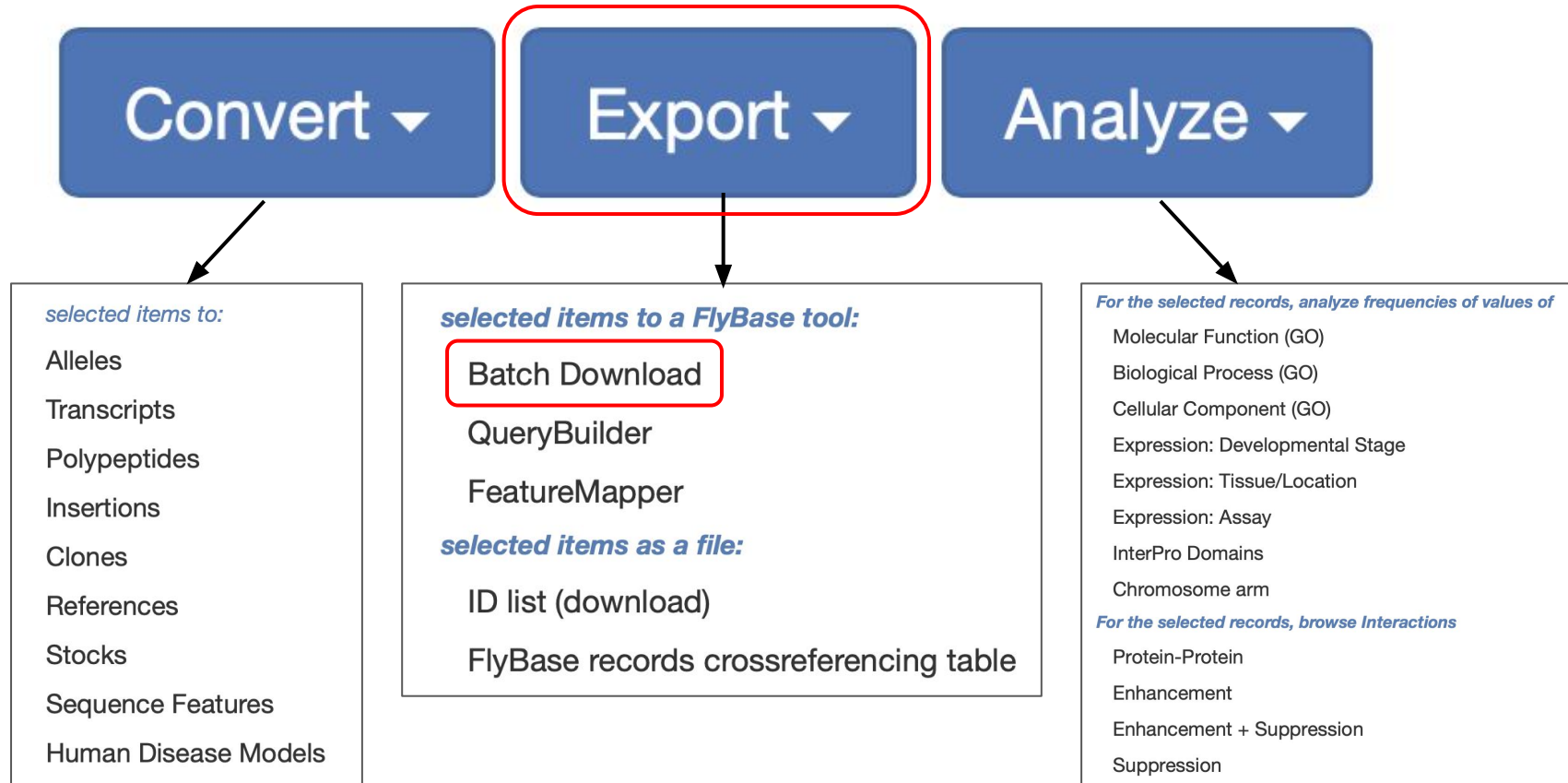
Items
1-7 of 7

Gene Results									
<input checked="" type="checkbox"/>	Symbol	Name	Annotation ID	Cytology	Scaffold	# Alleles	# Stocks	# Refs	Genome View
<input checked="" type="checkbox"/>	ItgaPS4	Integrin alphaPS4 subunit	CG16827	51E11-51E11	2R	7	6	64	GB JB
<input checked="" type="checkbox"/>	ItgaPS5	Integrin alphaPS5 subunit	CG5372	59E4-59E4	2R	6	5	54	GB JB
<input checked="" type="checkbox"/>	Itgbn	Integrin betanu subunit	CG1762	39A1-39A1	2L	18	9	94	GB JB
<input checked="" type="checkbox"/>	mys	myspheroid	CG1560	7D5-7D5	X	159	22	613	GB JB
<input checked="" type="checkbox"/>	mew	multiple edematous wings	CG1771	11E3-11E8	X	45	28	253	GB JB
<input checked="" type="checkbox"/>	if	inflated	CG9623	15A5-15A7	X	80	15	337	GB JB
<input checked="" type="checkbox"/>	scb	scab	CG8095	51E10-51E11	2R	48	22	218	GB JB

[← Previous](#) [1](#) [Next →](#)

Items
1-7 of 7

Hitlists - Tools



Batch Download

Batch Download

Help

Batch Download retrieves tabular data from FlyBase reports or precomputed files, for one or more IDs *etc.* you provide.
Looking for sequence data? Use our [Sequence Downloader](#) tool.

Enter IDs or Symbols:

FBgn0001250
FBgn0004456
FBgn0004657
FBgn0010395

or Upload File of IDs:

Browse

Reset form

Data source:

Report fields

Send results to:

Browser

We recommend splitting long lists of IDs into batches of not more than 1000 per submission. Longer lists may not complete successfully.

Continue to Select Fields

✓ Report fields

Precomputed files

✓ Browser

File

Batch Download

Select Fields

Check All

Uncheck All

Get Field Data

General Information

Check Section

Uncheck Section

- Symbol
- Name
- Feature Type
- Gene Model Status
- Gene Snapshot

Species Information

- Genus
- Species
- Abbreviation
- Annotation Symbol
- FlyBase ID

Genomic Location

Please see the "Genomic Location and Detailed Mapping Data" section for Cytogenetic map and Sequence Location information.

If you want sequence data instead: [Get Sequence](#)

Tag or Foreign Gene Data

- Tag or Foreign Gene Data

Families, Domains and Molecular Function

Check Section

Uncheck Section

- UniProt Protein Family
- UniProt Protein Domains

Please see "Gene Model and Products" -> "Polypeptide Data" for InterPro Domains

Please see "Gene Ontology: Function, Process & Cellular Component" for "GO Molecular Function".

Gene Ontology (GO): Molecular Function, Biological Process and Cellular Component

Check Section

Uncheck Section

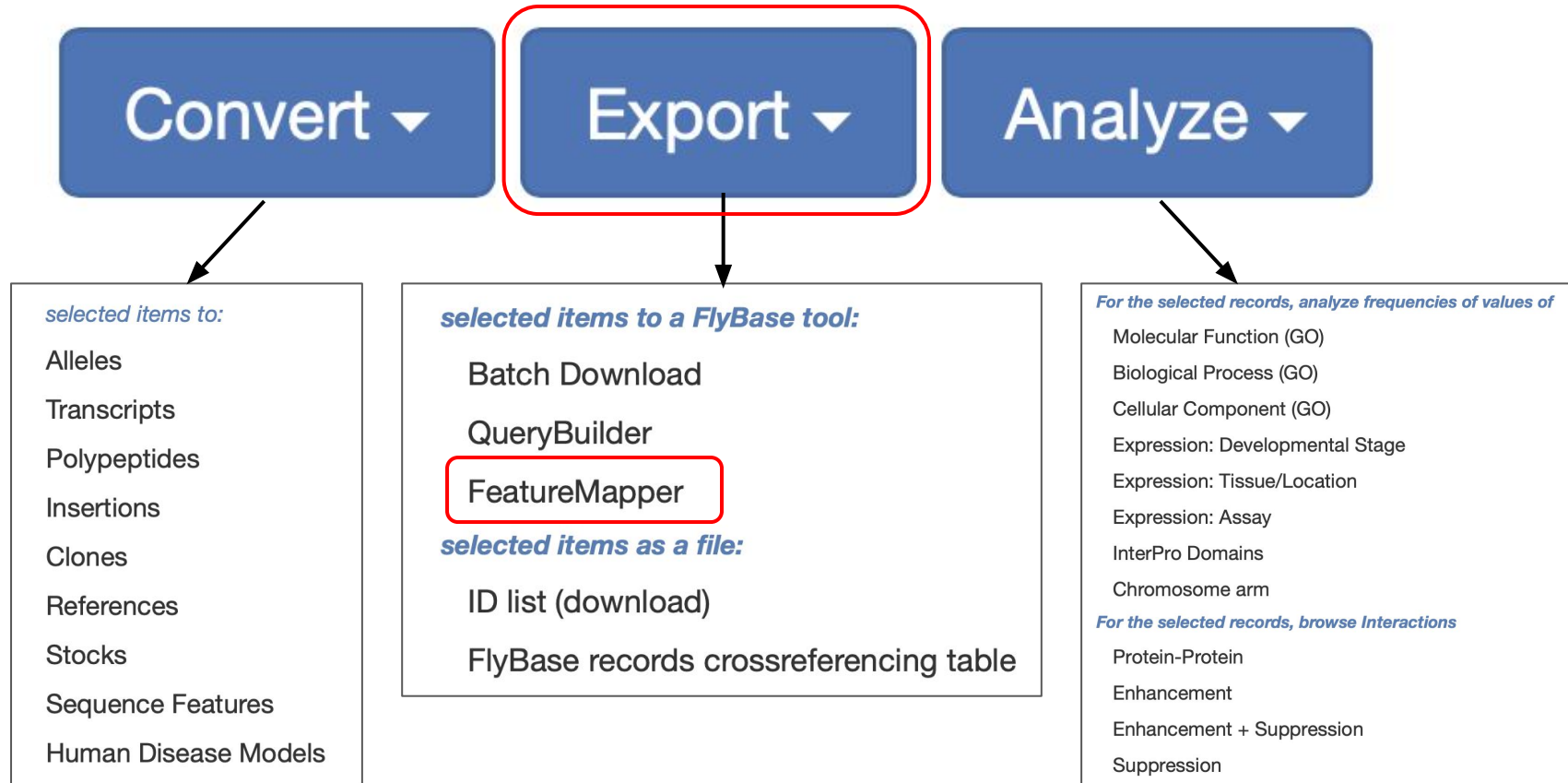
- Molecular Function
- Biological Process
- Cellular Component

Batch Download

[Download as a TSV file](#)

SUBMITTED ID	ANNOTATION SYMBOL	GENE SNAPSHOT	NAME	SYMBOL
FBgn0001250	CG9623	Inflated is one of 5 fly integrin alpha subunits, and when in a heterodimer with an integrin beta subunit makes a receptor for extracellular matrix proteins containing an RGD motif, such as Tig, wb laminin and Tsp. It is required in numerous tissues for cell migration and adhesion between cell layers. [Date last reviewed: 2016-12-01]	inflated	if
FBgn0004456	CG1771	Multiple edematous wings (Mew) is one of 5 fly integrin alpha subunits, and when in a heterodimer with an integrin beta subunit makes a receptor for the extracellular matrix protein laminin. Mew is required in numerous tissues for cell migration, adhesion between cell layers, and cell differentiation. [Date last reviewed: 2016-12-15]	multiple edematous wings	mew
FBgn0004657	CG1560	Myospheroid (Mys) is a β subunit of the integrin dimer. Integrin transmembrane receptors function as a link for the extracellular matrix and the intracellular actin cytoskeleton. Mys acts as adhesion/signaling protein regulating cellular adhesion, migration and survival. [Date last reviewed: 2016-06-30]	myospheroid	mys
FBgn0010395	CG1762	-	Integrin betanu subunit	Itgbn
FBgn0034005	CG16827	Insufficient genetic data for FlyBase to solicit a summary. [Date last reviewed: 2016-06-30]	Integrin alphaPS4 subunit	ItgaPS4
FBgn0034880	CG5372	Insufficient genetic data for FlyBase to solicit a summary. [Date last reviewed: 2016-06-30]	Integrin alphaPS5 subunit	ItgaPS5
FBgn0286785	CG8095	Scab (Scb) is the α -PS3 Integrin, believed to dimerise with mys and possibly Itgbn. Scb regulates adhesion, signalling, polarity and cell migration. Genetic interaction with Laminin genes implicates them as extracellular matrix ligands. [Date last reviewed: 2016-06-23]	scab	scb

Hitlists - Tools



Export to FeatureMapper

Mapping Options

Reference Landmark(s) or Region(s): [Feature Mapper help](#)

FBgn0001250
FBgn0004456
FBgn0004657
FBgn0010395
FBgn0034005
FBgn0034880
FBgn0286785

Region type to map: Sequence of the Landmark

Include overlapping (not fully enclosed within query region) features

Species: D. melanogaster

Group output features by type

Gene Models check all

- Gene
- mRNA (transcript)
- exon
- 5' UTR
- 3' UTR
- tRNA
- miRNA
- snRNA
- snoRNA
- CDS (polypeptide)
- Natural TE

Aligned Evidence check all

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

Microarray Features check all

- Affymetrix v1
- Affymetrix v2

Noncoding Features check all

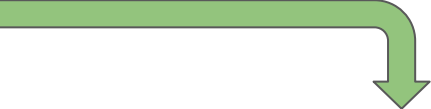
- Regulatory Regions
- Insulator class I
- Insulator class II
- Protein binding site
- Enhancers
- Silencers
- TFBS - HOT spot analysis
- TFBS - zinc finger domain
- TFBS - homeodomain
- TFBS - helix-loop-helix domain
- TFBS - BTB/POZ domain
- TFBS - mesoderm
- TFBS - other
- Origin of replication
- Histone Modifications - mesoderm
- RNA Editing Sites
- Putative Brain Enhancers
- VDRC Vienna Tiles GAL4 lines

Mapped Mutations check all

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

RNAi Reagents and Data check all

- 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



FBgn0004456

All features [Get combined GFF file](#)

VDRC RNAi amplicons [to HitList](#)

	X:13,217,327..13,217,918	-->	RNAi_reagent	dsRNA-KK101081
	X:13,236,169..13,236,438	-->	RNAi_reagent	dsRNA-KK110893
	X:13,217,972..13,218,300	<--	RNAi_reagent	dsRNA-GD1230
	X:13,236,171..13,236,471	<--	RNAi_reagent	dsRNA-GD15842
	X:13,257,724..13,258,044	<--	RNAi_reagent	dsRNA-GD9103

TRiP RNAi amplicons [to HitList](#)

	X:13,218,213..13,218,233	<--	RNAi_reagent	dsRNA-HMS02849
	X:13,217,327..13,217,918	<--	RNAi_reagent	dsRNA-JF02694

FBgn0034005

All features [Get combined GFF file](#)

VDRC RNAi amplicons [to HitList](#)

	2R:15,261,827..15,262,174	-->	RNAi_reagent	dsRNA-KK105806
	2R:15,260,688..15,260,941	-->	RNAi_reagent	dsRNA-GD2007

TRiP RNAi amplicons [to HitList](#)

	2R:15,261,760..15,261,780	-->	RNAi_reagent	dsRNA-HMC02928
	2R:15,259,308..15,259,722	-->	RNAi_reagent	dsRNA-HM05021

FBgn0010395

All features [Get combined GFF file](#)

VDRC RNAi amplicons [to HitList](#)

	2L:21,055,178..21,055,423	-->	RNAi_reagent	dsRNA-GD7
	2L:21,055,178..21,055,423	-->	RNAi_reagent	dsRNA-GD2503
	2L:21,056,408..21,056,764	-->	RNAi_reagent	dsRNA-GD14392

TRiP RNAi amplicons [to HitList](#)

	2L:21,053,490..21,053,510	-->	RNAi_reagent	dsRNA-HMJ23472
	2L:21,056,320..21,056,883	-->	RNAi_reagent	dsRNA-HM05089

Sequence Downloader

FlyBase Sequence Downloader

[Help](#)

Mode

ID

FlyBase ID

FBgn0000490

Type

Gene Region

[View Sequence](#)

- ✓ ID
- Bulk ID
- Bulk Region

- ✓ Gene Region
- Extended Gene Region (2kb 5' and 3')
- CDS
- Introns
- Exons
- Translations
- Transcripts
- 5' UTR
- 3' UTR
- Clones
- Sequence Features
- Recombinant Constructs

Sequence Downloader

FlyBase Sequence Downloader

Help

Mode

Bulk ID

FlyBase IDs

FBgn0001250
FBgn0004456
FBgn0004657
FBgn0010395
FBgn0034005
FBgn0034880
FBgn0286785

Sequence Type

5' UTR

Output to

File

Download



```
>FBtr0073732 type=five_prime_untranslated_region;
loc=X:complement(13256832..13257183);MD5=747a1d6a31a1b500a21
3404b04595087; length=352; parent=FBgn0004456;
release=r6.25; species=Dmel;
CATTTTATTTTTTCGTTTGCCCGTCGAAGAAGGCGAAGCGAAGAAAAATATCTGCCGCTT
CGAACGCACGCTATCAATGTCTAATCTTTACGTAATATCATTTTTTTTTTGGTGTGTG
GACTGAAACGAAGAAATATACGAGCAAAGTGAAAATACGTGGACTACCTGTTGATCCA
AAGATCTCAACGGATTTAGTGTGTGTGTGTGTGTGCGGTGCGGTGTTTTTTTTTCGGTTC
TTGTTTTTGTTCGAATAGTGCAAAGGAAAAAGTAATCACTTTGCCGTGTTCCGTTGCACGT
GCACGAAATATTTATAAAATATATATATAAAAAAAAAAACATAAAATTAACG
>FBtr0087369 type=five_prime_untranslated_region;
loc=2R:15248784..15249295;MD5=34d02938d0bf72114791979d9c5498
6f; length=512; parent=FBgn0286785; release=r6.25;
species=Dmel;
ATTTCCGGTTCGGGTCGTCGAAGAGTGAACGTCACGACTGATCTGTTGCCCGCTTAT
GGACTAGCGTTGCCGAAATAAAACCCATATATGCTATAGGCTATATGCTATATGCTATT
TATTAGCTATTAGCTACACATCAGTGCAGAGATACGGCGCGTGCAGTAGTCCGATTAAC
AACGACTTGTCCCGCAAAGTGACAAAGGCCAAAACAAGTTGCCACCTAGATCTCCG
GAGAGATTTGCGAAAGCCCGCGCTGGCTTAATTAATTAATATAAATTTCCAGAGAACG
CCGCCCATACAGCAGCGATCTCGAGTGTGTGAGTGTCTCAATTAATTCRAATAATCA
TAAATAAGTGAGTCCAACGGATCCTGTTTTCAGTGTGTGCTCTCGGCCAATTTGTA
TTTTATTGCCCTCAGAAGAGCCGGGTCATTCAAGTGAATTACCAAAATTCAGAAATTA
CGAATCGCAAACTTAATTTTGATGAGTTCCA
>FBtr0071105 type=five_prime_untranslated_region;
loc=X:join(8061645..8061725,8063840..8063957);MD5=d28be4483e
39eade63219439aa67d7f5; length=199; parent=FBgn0004657;
release=r6.25; species=Dmel;
ACATTGACTGTTGTCCACCCCTGCTGACACCATCTGGTAACACTGAACAACAGACCAG
ATCCGGGCAAAGCAAAGCACCCGGCCAGAGAAGCGACAGACAGAGACACACACGCA
CCTGGCACATCCAAGGATATCCGATCTCTGCATCCCGAATCATCCACACCTGCACCA
ACCGCTAACGCCAAAGCC
>FBtr0301353 type=five_prime_untranslated_region;
loc=X:complement(join(16780057..16780196,16782470..16782545)
);MD5=2253be4a2f02a26f56555d0802e2faa2; length=216;
parent=FBgn0001250; release=r6.25; species=Dmel;
GGCGGTGCTAAGATGATCTTACAGATACTGTGCGCCATTAGTCTGTTAAATCTTGC
TGCCTTCGCTGCCGAGCTGTGAAGATGCAAGCGTAAAGTGTCTAAGTCTTGGATGCAAGC
CAACGAAATCGATATACAAATTCGAGAGAAGTAGAAAATACAGCAAAAAGTAAAGTAAT
CAGCTCACCGAATAAAGAAGTTAGAAGTCGACAAA
>FBtr0346711 type=five_prime_untranslated_region;
loc=2L:21053033..21053133;MD5=06e564488126816dafc3c01484301e
9c; length=101; parent=FBgn0010395; release=r6.25;
species=Dmel;
ATATTAGTTACGCTGGACTTCTAAGGCCGAACGGTCGCTAGCTCCCTTAGCAAAAAAAT
ATATGTGATCSTGGTAGTGAAAATTCACGCAATTTTGAAA
```


Part 3: Bulk data files (FTP site)

- Overview
- FTP site & Web interface
- Examples
- Use in Batch Download
- Survey

Bulk data files

- >100 bulk data files available
- Each contain defined subsets of data in different file formats:
 - TSV
 - FASTA, GFF, GTF
 - Chado XML
- Most generated with each FlyBase release, archived versions also available



Overview

Current Release

Archived Releases

Map Conversion

FTP Site:

Releases (FTP)

Genomes (FTP)

FlyBase:Downloads Overview


Contents [hide]

- 1 Introduction
 - 1.1 Opening compressed files
 - 1.2 [Archived Data](#)
- 2 Main Data Set
 - 2.1 Postgres Chado Database Dump
 - 2.2 Drosophila Data
- 3 Bulk data files
 - 3.1 Synonyms
 - 3.1.1 FlyBase Synonyms (fb_synonym_*.tsv)
 - 3.2 Genes
 - 3.2.1 Genes data (Chado XML)
 - 3.2.2 Genetic interaction table (gene_genetic_interactions_*.tsv)
 - 3.2.3 RNA-Seq RPKM values (gene_rpk_report_fb_*.tsv.gz)
 - 3.2.4 Physical interaction table (physical_interactions_fb_*.tsv.gz)
 - 3.2.5 Physical interaction MITAB file (physical_interactions_mitab_fb_*.tsv.gz)
 - 3.2.6 Functional complementation table (gene_functional_complementation_*.tsv)
 - 3.2.7 FBgn <=> DB Accession IDs (fbgn_NAseq_Uniprot_*.tsv)
 - 3.2.8 FBgn <=> Annotation ID (fbgn_annotation_ID_*.tsv)

FTP site and webpage interface

Index of ftp://ftp.flybase.net/releases/current/

 [Up to higher level directory](#)

Name	Size	Last Modified
File: FB2018_06		21/12/2018 18:07:00 GMT
README		21/12/2018 20:32:00 GMT
 chado-xml		21/12/2018 20:51:00 GMT
 collaborators		21/12/2018 18:10:00 GMT
dana_r1.06		21/12/2018 18:24:00 GMT
dere_r1.05		21/12/2018 18:25:00 GMT
dgri_r1.05		21/12/2018 18:25:00 GMT
dmeI_r6.25		21/12/2018 18:25:00 GMT
dmoj_r1.04		21/12/2018 18:26:00 GMT
dper_r1.3		21/12/2018 18:26:00 GMT
dpse_r3.04		21/12/2018 18:24:00 GMT
dsec_r1.3		21/12/2018 18:27:00 GMT
dsim_r2.02		21/12/2018 18:24:00 GMT
dvir_r1.07		21/12/2018 18:24:00 GMT
dwil_r1.05		21/12/2018 18:28:00 GMT
dyak_r1.05		21/12/2018 18:28:00 GMT
 precomputed_files		21/12/2018 18:10:00 GMT
 psql		21/12/2018 20:53:00 GMT

 Main Data Set		
 Synonyms		
 Genes		
Item	Chado XML Other	
Genes data	download	
Genetic Interaction table		gene_genetic_interactions_fb_2018_06.tsv.gz
RNA-Seq RPKM values		gene_rpkm_report_fb_2018_06.tsv.gz
Physical Interaction table	TSV	physical_interactions_fb_2018_06.tsv.gz,
	MITAB TSV	physical_interactions_mitab_fb_2018_06.tsv.gz
Functional complementation table		gene_functional_complementation_fb_2018_06.tsv.gz
FBgn ↔ DB Accession IDs		fbgn_NAseq_Uniprot_fb_2018_06.tsv.gz
FBgn ↔ Annotation ID		fbgn_annotation_ID_fb_2018_06.tsv.gz
FBgn ↔ GLEANR IDs		fbgn_gleanr_fb_2018_06.tsv.gz
FBgn ↔ FBtr ↔ FBpp IDs		fbgn_fbtr_fbpp_fb_2018_06.tsv.gz
FBgn Exons ↔ Affy1		fbgn_exons2affy1_overlaps.tsv.gz
FBgn Exons ↔ Affy2		fbgn_exons2affy2_overlaps.tsv.gz
Genes GO data		gene_association.fb.gz
Genes map table		gene_map_table_fb_2018_06.tsv.gz
Automated gene summaries		automated_gene_summaries.tsv.gz
Gene snapshots		gene_snapshots_fb_2018_06.tsv.gz
Drosophila Orthologs		dmel_orthologs_in_drosophila_species_fb_2018_06.tsv.gz
Unique Protein Isoforms		dmel_unique_protein_isoforms_fb_2018_06.tsv.gz
Non-coding RNA genes (TSV)		ncRNA_genes_fb_2018_06.tsv.gz
Non-coding RNAs (JSON)		ncRNA_genes_fb_2018_06.json.gz
 Gene Groups		
 Alleles and Stocks		
 Orthologs		
 Human Disease		
 Nomenclature		
 Ontology Terms		
 Genomes: Annotation and Sequence		
 Transcripts and Polypeptides		
 Transposons, Transgenic Constructs, and Insertions		
 Aberrations		
 Large Dataset Metadata		
 Clones		
 References		
 Drosophila Researchers		
 Map Conversion Tables		

Bulk data file - examples

fbgn_fbtr_fbpp_fb_2018_06.tsv.gz

```
## FlyBase current_gene_product report
## Generated: Fri Nov 16 23:02:35 2018
## Using datasource: dbi:Pg:dbname=fb_2018_06_reporting;host=flysql15;port=5432...

##FlyBase_FBgn  FlyBase_FBtr  FlyBase_FBpp

FBgn0095581    FBtr0123263    FBpp0121755
FBgn0095581    FBtr0389002    FBpp0348660
FBgn0081305    FBtr0279597    FBpp0278035
FBgn0256311    FBtr0296469
```

dme1_human_orthologs_disease_fb_2018_06.tsv.gz

```
## FlyBase D. melanogaster-Human Orthologs and associated diseases report
## Generated: Fri Nov 16 23:44:39 2018
## Using datasource: dbi:Pg:dbname=fb_2018_06_reporting;host=flysql15;port=5432...

##Dme1_gene_ID  Dme1_gene_symbol  Human_gene_HGNC_ID  Human_gene_OMIM_ID  Human_gene_symbol  DI0PT_score  OMIM_PhenoType_IDs  OMIM_PhenoType_IDs[name]
FBgn0031081    Nep3              HGNC:14668          OMIM:618104         MMEL1              3              617017,617018       617017[CHARCOT-MARIE-TOOTH DISEASE, AXONAL, TYPE 2T; CMT2T],617018[SPINOCEREBELLAR ATAXIA 43; SCA43]
FBgn0031081    Nep3              HGNC:7154           OMIM:120520         MME                 3
FBgn0031081    Nep3              HGNC:13275         OMIM:610145         ECE2                8
```

dme1-all-tRNA-r6.25.fasta.gz

```
>FBtr0070001 type=tRNA; loc=X:20025099..20025170; ID=FBtr0070001; name=tRNA:Pro-CGG-1-1-RA; dbxref=FlyBase:FBtr0070001,FlyBase_Annotation_IDs:CR32826-RA,RNACentral:URS00000C18F2_7227; MD5=376f4e117637d6ac68f3a5d7195ff7be; length=72; parent=FBgn0052826; release=r6.25; species=Dme1;
GGCTCGTTGGTCTAGGGGTATGATTCTCGCTTCGGGTGCGAGAGGTCCCGGGTTCAAATCCCGGACGAGCCC
>FBtr0070533 type=tRNA; loc=X:complement(3427452..3427523); ID=FBtr0070533; name=tRNA:Gln-TTG-2-1-RA; dbxref=FlyBase:FBtr0070533,FlyBase_Annotation_IDs:CR32493-RA,RNACentral:URS0000360ADA_7227; MD5=565bec4ff3eda493904bab301eb48dfb; length=72; parent=FBgn0052493; release=r6.25; species=Dme1;
GGTTCATGTGGTAAACGGTTAGCACTCTGGACTTTGAATCCAGCGATCCGAGTTCAAATCTCGGTAGAACCT|
```

Using bulk data files in Batch Download

Batch Download

Help

Batch Download retrieves tabular data from FlyBase reports or precomputed files, for one or more IDs etc. you provide.

Looking for sequence data? Use our [Sequence Downloader tool](#).

Enter IDs or Symbols:

FBgn0001250
FBgn0004456
FBgn0004657
FBgn0010395

or Upload File of IDs:

Browse

Data source:

Precomputed files

Send results to:

File

We recommend splitting long lists of IDs into batches of not more than 1000 per submission. Longer lists may not complete successfully.

Reset form

Continue to Select Fields

Select precomputed file (only files relevant to DataSet fbgn are shown)

Alleles: fbal_to_fbgn_fb_2018_06.tsv
Genes: automated_gene_summaries.tsv
Genes: dmel_unique_protein_isoforms_fb_2018_06.tsv
Genes: fbgn_NAseq_Uniprot_fb_2018_06.tsv
Genes: fbgn_annotation_ID_fb_2018_06.tsv
Genes: fbgn_exons2affy1_overlaps.tsv
Genes: fbgn_exons2affy2_overlaps.tsv
Genes: fbgn_fbtr_fbpp_fb_2018_06.tsv
Genes: fbgn_gleanr_fb_2018_06.tsv
Interactions: gene_functional_complementation_fb_2018_06.tsv
Interactions: gene_genetic_interactions_fb_2018_06.tsv
Gene groups: gene_group_data_fb_2018_06.tsv
Genes: gene_map_table_fb_2018_06.tsv
Genes: gene_rpkm_report_fb_2018_06.tsv
Genes: gene_snapshots_fb_2018_06.tsv
Interactions: physical_interactions_fb_2018_06.tsv
GO: gene_association.fb
Orthologs: dmel_human_orthologs_disease_fb_2018_06.tsv
Orthologs: dmel_orthologs_in_drosophila_species_fb_2018_06.tsv
:Synonyms: fb_synonym_fb_2018_06.tsv

Retrieve data



Bulk data files: survey

<https://www.surveymonkey.co.uk/r/FlyBasebulkdatafiles>



FlyBase bulk data files

Introduction

FlyBase provides many (>100) downloadable bulk files, each containing defined subsets of data in different formats. These files are accessed via the 'Downloads' menu of the main navigation bar on any FlyBase page and can be obtained via a [FlyBase-styled webpage interface](#) (left panel in figure below) or standard [ftp site interface](#) (right panel). Details of the contents and format of each file are provided in [this document](#), which is available via the 'Overview' link under the 'Downloads' menu on the navigation bar or by clicking the 'Help' button at the top of the webpage interface. Please take a few minutes to browse these pages before completing the survey.

Part 4: Direct database queries & APIs

- Query public database via an SQL client
- Download entire postgresSQL database
- APIs

Downloading the Postgres database (100 GB!)

Index of ftp://ftp.flybase.net/releases/current/psql/



[Up to higher level directory](#)

Name	Size	Last Modified	
File: FB2018_06.sql.gz.00	1945600 KB	21/12/2018	12:35:00 GMT
File: FB2018_06.sql.gz.01	1945600 KB	21/12/2018	12:36:00 GMT
File: FB2018_06.sql.gz.02	1945600 KB	21/12/2018	12:37:00 GMT
File: FB2018_06.sql.gz.03	1945600 KB	21/12/2018	12:37:00 GMT
File: FB2018_06.sql.gz.04	1945600 KB	21/12/2018	12:38:00 GMT
File: FB2018_06.sql.gz.05	1202090 KB	21/12/2018	12:38:00 GMT
File: README	2 KB	21/12/2018	18:10:00 GMT
File: md5sum.txt	1 KB	21/12/2018	18:10:00 GMT

Downloading chado XML files

Index of ftp://ftp.flybase.net/releases/current/chado-xml/



[Up to higher level directory](#)

Name	Size	Last Modified	
File: chado_FBab.dtd	4 KB	21/11/2018	21:31:00 GMT
File: chado_FBab.xml.gz	24696 KB	20/11/2018	21:45:00 GMT
File: chado_FBal.dtd	4 KB	21/11/2018	21:32:00 GMT
File: chado_FBal.xml.gz	209137 KB	20/11/2018	22:39:00 GMT
File: chado_FBba.dtd	3 KB	21/11/2018	21:30:00 GMT
File: chado_FBba.xml.gz	997 KB	20/11/2018	22:39:00 GMT
File: chado_FBcl.dtd	4 KB	21/11/2018	21:35:00 GMT
File: chado_FBcl.xml.gz	415303 KB	21/11/2018	00:53:00 GMT
File: chado_FBgg.dtd	3 KB	21/11/2018	21:30:00 GMT
File: chado_FBgg.xml.gz	4698 KB	21/11/2018	00:55:00 GMT
File: chado_FBgn.dtd	7 KB	19/12/2018	05:22:00 GMT
File: chado_FBgn.xml.gz	1251093 KB	20/12/2018	09:59:00 GMT

Connecting to the public database

If you have a PostgreSQL client application installed, you can connect to the FlyBase public read only Chado database:

```
$ psql -h chado.flybase.org -U flybase flybase
```

(These instructions are included in the Downloads -> Overview document)

API access: <http://api.flybase.org/api/v1.0>

Data Endpoint	Retrieves:
HitList	HitList objects for a given FlyBase object(s)
Updates	The recent update information for a given FlyBase object
Gene Summaries	Automatically generated summary for a given gene
Protein Domains	All domains for a given gene/protein
Datasets	Features associated with a FlyBase dataset object
Sequence	Sequence data by genomic location or FBid
Chado	ChadoXML data for a given FlyBase object

Please let us know what other API endpoints would be useful

FlyBase.org API

1.0 OAS3

API access to various FlyBase.org data sets

[the developer - Website](#)

[Apache 2.0](#)

Servers

[http://api.flybase.org/api/v1.0/ - FlyBase API](http://api.flybase.org/api/v1.0/) ▾

Domains Endpoints for retrieving protein domain information

GET `/domain/{flyBaseGeneId}` Gets all domains for a given gene.

GET `/domain/{flyBasePolypeptideId}` Gets all domains for a given polypeptide.

Chado Endpoints for retrieving ChadoXML

GET `/chadoxml/{flyBaseId}` Fetch ChadoXML for a FlyBase object by ID.

Datasets Endpoints for retrieving FlyBase Dataset objects.

GET `/downloads/FB1c/associated_data/{fb1cId}` Fetch associated data for FlyBase Datasets.

Fetches lists of strains, cell lines, and features that are associated with a FlyBase dataset object.

<http://api.flybase.org/api/v1.0/domain/FBgn0001250>

```
{
  "resultset": {
    "api_version": "1.0",
    "data_version": "FB2018_06",
    "query_url": "http://api.flybase.org/api/v1.0/domain/FBgn0001250",
    "query_time": "2019-01-25T14:27:33",
    "result": [
      {
        "proteins": [
          {
            "length": 1396,
            "regions": [
              {
                "source": "dmel_domain_PFAM",
                "abbrev": "FG-GAP",
                "name": "FG-GAP repeat",
                "end": 374,
                "pfam": "PF01839",
                "interpro": "IPR013517",
                "classification": "Repeat",
                "start": 332
              },
            ],
          },
        ],
      },
    ],
  },
}
```

Part 5: Further information

FlyBase Community

Fast-Track Your Paper

**FlyBase Community
Advisory Group**

Gene Snapshots

FlyBase Forum

Find A Person

Newsletter

bionet.dros

FlyGene Wiki

Twitter

Release Notes

New In This Release

Release Schedule

FlyBase Positions

Citing FlyBase

FlyBase Consortium

FlyBase Publications

FlyBase Presentations

FlyBase Support

Video Tutorials 

Tool help

Report help

Info for Authors ▶

Linking to/from FlyBase

Nomenclature

Curation documentation ▶

New to Flies

Help Index

Contact FlyBase

Summary

1. Obtaining data on single genes (web)
2. Obtaining data on multiple genes (web)
3. Bulk data files (FTP site)
4. Database downloads, queries & API access
5. Additional information/help
6. Questions / feedback / use cases