Regulatory Element Database for Fly

Subjects: Biology

Contributor: Soile V. E. Keränen , Angel Villahoz-Baleta , Andrew E. Bruno , Marc S. Halfon

The REDfly (Regulatory Element Database for Fly) database (http://redfly.ccr.buffalo.edu/, accessed on 27 July 2022) integrates all of the available insect cis-regulatory information from multiple sources to provide a comprehensive collection of known regulatory elements.

insects Drosophila regulatory genomics gene regulation cis-regulatory module enhancer genome annotation

1. Introduction

The turn-of-the-century advent of fully sequenced metazoan genomes brought with it the first genome annotations, which were largely confined to positions of confirmed and predicted genes, and typically housed in community-specific model-organism databases, e.g., [1][2][3][4]. Remarkably, over two decades later, the major databases (see [5]) are still mostly lacking annotation of non-coding regulatory sequences. These sequences include distal "*cis*-regulatory modules" (CRMs), a generic term encompassing such regulatory elements as enhancers, which mediate positive gene regulation; silencers, involved in negative regulation; and a growing number of additional elements that are not easily classified including PREs, super-enhancers, insulators, tethering elements, and others [6][7][8][9][10][11].

2. Utility of REDfly

Prior to the development of REDfly, large-scale analyses of regulatory sequences were challenging to conduct, as the bulk of the existing regulatory data was distributed among hundreds of individual publications. Consequently, what few analyses were completed were performed on small and frequently biased sets of CRMs, such as a limited subset of early developmental pair-rule stripe enhancers in *Drosophila*, e.g., [12][13][14]. By curating these data and making them findable, accessible, interoperable, and usable (FAIR) [15], REDfly made it possible to bring statistical, computational, and comparative genomics methods to bear on their study. REDfly enabled the first-ever large-scale, relatively unbiased analysis of CRMs, which immediately revealed novel insights into CRM-sequence composition, differences among tissue-specific groups of CRMs, and an early indication of the presence of enhancer RNAs (eRNAs) as a prevalent CRM characteristic [16]. REDfly, by continuing to compile the data from hundreds and eventually thousands of individual experiments scattered throughout four decades of literature, subsequently proved instrumental in facilitating studies in a wide variety of research areas, including:

Biology of CRMs. REDfly has been used to investigate the organization of TFBSs within CRMs [17] and how combinatorial binding influences CRM activity [18]. Soluri et al. [19] investigated how pioneer TFs control chromatin accessibility, and Blick et al. [20] examined the ability of CRMs to act in *trans*. REDfly data helped to illustrate how CRMs can have multiple functions [21], such as dual use as both enhancers and Polycomb response elements [22], or as both enhancers and silencers [23].

Interpretation of genomic data. REDfly has been critical for interpreting data from large-scale genomics projects including TF binding studies, e.g., [24][25] and studies of insulators [26][27]. A study challenging the understanding of which epigenetic marks characterize regulatory sequences depended on REDfly data [28]. REDfly has been used to study chromosome domains and chromatin "states", e.g., [29][30][31][32], to explore 3D-chromatin conformation [33][34], to study ncRNA and eRNA expression [35][36], and to validate scATAC-seq approaches, e.g., [37][38].

Computational CRM discovery. REDfly has played a dramatic role in methods for computational CRM discovery, both as a source of training data and as a method for validating predictions, e.g., [39][40][41][42][43][44][45][46][47]. Su et al. [48] used REDfly data to assess CRM-discovery approaches, which would have been impossible without REDfly. Computational CRM-discovery methods using REDfly for training data also can identify CRMs in diverse insect species [49] and, as such, provide a powerful tool for annotating insect regulatory genomes [50].

CRM evolution. REDfly has enabled studies of CRM evolution and TFBS turnover, e.g., [51][52][53][54][55][56][57]. Wang et al. [58] used REDfly data to investigate the selective pressure on DNA shape at TF binding sites, and Peng et al. [59] explored the relationship between chromatin accessibility and TF binding to predict evolutionary changes in enhancer activity.

As can be seen from these examples, REDfly is an important source of raw data for analysis, hypothesis generation, assessment, validation, and empirical research.

3. REDfly Data Model

REDfly curates two types of data: CRMs and transcription factor binding sites (TFBSs), with CRMs being the main focus. Historically, CRM annotations have been drawn from reporter gene assays in transgenic animals or cultured cells, but an increasingly diverse set of assays are starting to be included. In particular, several years ago REDfly started to capture CRMs identified through various "X-seq" assays, such as ATAC-seq, FAIRE-seq, DNase-seq, ChIP-seq, etc., as well as from purely computational predictions, in recognition of the fact that many regulatory sequences are presently being defined by these methods. There is considerable debate in the regulatory genomics field as to just how CRMs should be defined, with several studies indicating that the different methods of CRM identification have led to widely non-overlapping results and raising questions as to which, if any, of these methods most accurately identify CRMs [60][61][62]. As a result, REDfly separates its CRM data into four distinct subclasses: *reporter constructs (RC)*, *CRM_segments*, *predicted CRMs (pCRM)*, and *inferred CRMs (iCRM)*. *RCs* and *CRM_segments* are drawn from activity-based assays (**Figure 1**, left), primarily gene-expression data from either reporter genes or from native genes following mutation or deletion of regulatory sequences. *RCs* mainly represent

reporter-gene results, assayed either in transgenic animals or in cell-culture assays; the two types of assays can be independently searched. The *CRM_segment* class contains sequences that are demonstrated to be necessary for gene regulation but, unlike in a reporter gene assay, are not necessarily sufficient. Such sequences can be obtained from the analysis of small chromosomal deletions or site-directed mutagenesis but are increasingly being found in the literature as a result of CRISPR/Cas9-mediated targeted sequence deletions. *pCRMs*, on the other hand, reflect CRMs identified by assays that do not require demonstration of activity, for example, the presence of histone modifications, or computational predictions (**Figure 1**, right). *iCRMs* are not curated from the literature, but represent putative regulatory elements based on the analysis of other REDfly data (see below).

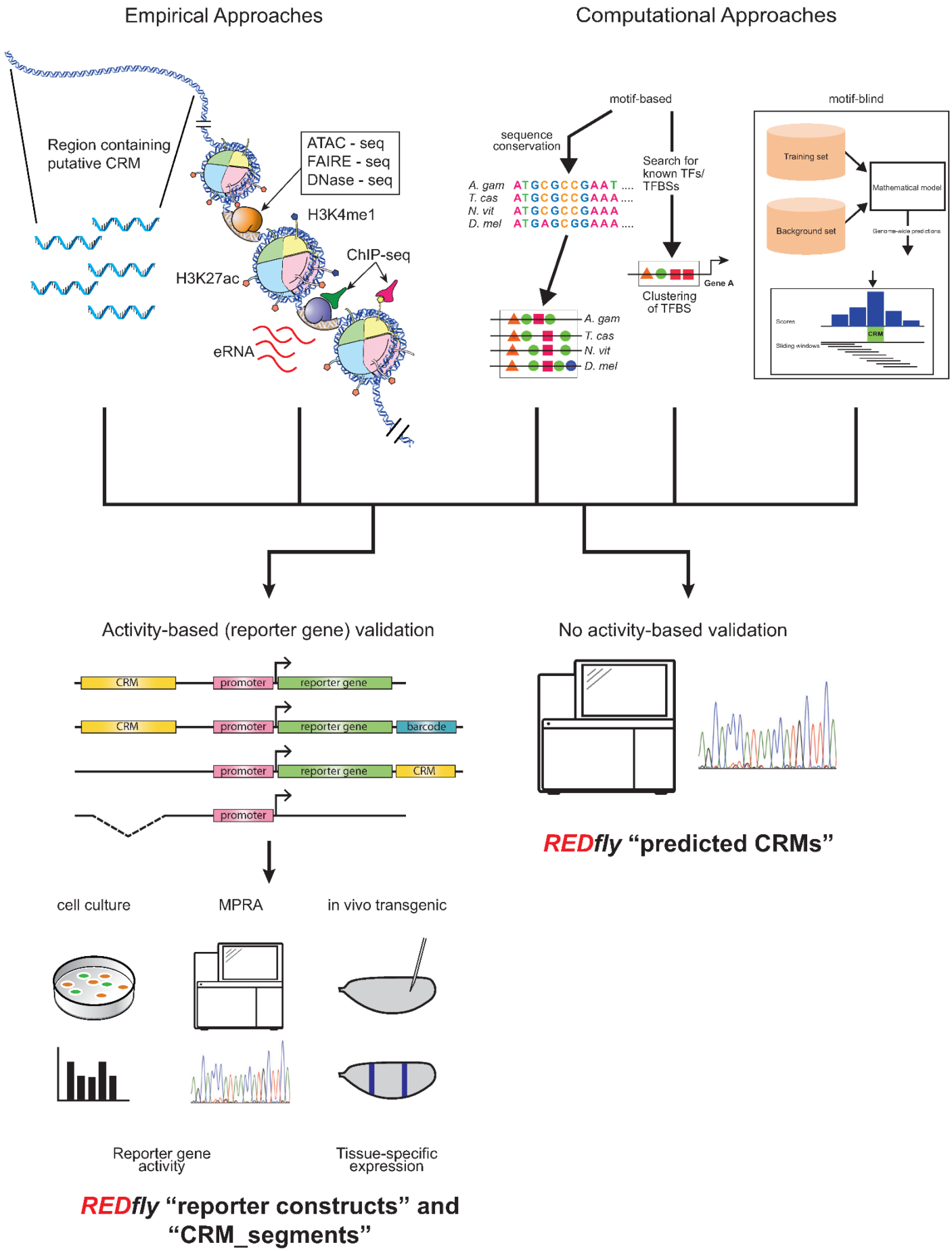
Empirical Approaches
Computational Approaches
Region containing putative CRM
ATAC - seq
FAIRE - seq
DNase - seq
H3K4me1
H3K27ac
ChIP-seq
eRNA
motif-based
sequence conservation
A. gam ATGCGCCGAAT ....
T. cas ATGCGCCGAAA ....
N. vit ATGCGCCGAAA
D. mel ATGAGCGGAAA ....
Search for known TFs/ TFBSs
Gene A
Clustering of TFBS
A. gam
T. cas
N. vit
D. mel
motif-blind
Training set
Background set
Mathematical model
Genome-wide predictions
Scores
CRM
Sliding windows
Activity-based (reporter gene) validation
No activity-based validation
CRM
promoter
reporter gene
barcode
cell culture
MPRA
in vivo transgenic
Reporter gene activity
Tissue-specific expression
REDfly "predicted CRMs"
REDfly "reporter constructs" and "CRM_segments"

Figure 1. Activity-based and non-activity-based methods for defining regulatory sequences. Top: A wide variety of methods exist for identifying regulatory sequences based on both empirical (**left**) and computational (**right**) approaches. Empirical approaches include unbiased testing of non-coding DNA regions as well as selection of sequences based on chromatin accessibility, histone post-translational modification, transcription factor binding, production of enhancer RNAs, and others. Computational approaches may include assessment of sequence conservation, presence of transcription factor binding motifs, or various machine-learning methods. Bottom: Results from these regulatory element discovery methods can be obtained with (**left**) or without (**right**) the use of activity-based criteria. Activity-based criteria typically involve some sort of reporter-gene assay, which might be performed in cultured cells, using next-generation sequencing in a "massively parallel reporter assay" (MPRA), or in transgenic animals; recently, testing via genomic deletion via CRISPR/Cas9 has also been gaining popularity. REDfly classifies regulatory sequences derived from these methods as *reporter constructs (RCs)* and *CRM_segments*, while any methods that identify regulatory sequences without recourse to activity-based criteria are referred to as *predicted CRMs (pCRMs)*. These somewhat historical definitions should not be construed to imply that one or the other type of data is more accurate or "correct".

4. Species in REDfly

Although REDfly has historically been focused on *Drosophila melanogaster*, the clear value of comparative genomics and of working with non-traditional model organisms, a vast increase in the number of sequenced insect genomes, and the small but growing availability of both predicted and validated insect regulatory sequences has led us to expand REDfly by implementing the ability to curate regulatory sequences for additional insect species. Information on which species are represented can be found on the "Species" page; current species include the mosquitoes *Anopheles gambiae* and *Aedes aegypti* and the beetle *Tribolium castaneum*. Additional species will be added as data accumulate. Since insect CRMs are often tested using transgenic *Drosophila*, REDfly divides the sequence and gene-feature data and the expression pattern and cell-line data into separate components. Each REDfly record has both a "sequence from" and an "assayed in" component. Sequence and gene-feature data are linked to the former, and anatomy and staging data to the latter. While it is preferable to describe species-specific reporter-gene-expression patterns using the proper species-specific anatomy ontology, many species lack an ontology as rich in terms as that for *Drosophila*. Therefore, terms from the *Drosophila* ontology can also be used to annotate expression in other species.

In order to facilitate research using these newly added genomes, the researchers have implemented interfaces for *BLAT* [63] and in silico *PCR* [64] for each species included in REDfly. These can be accessed through the "Species" page.

5. Contents of REDfly

REDfly has continued to expand its contents at a rapid rate (**Table 1**). Since the end of 2019, the number of curated publications has increased by 30%, leading to an increase in the total number of Reporter Construct

records by over 25% and in the number of pCRMs by almost 60%. Not reflected in these numbers, however, is an ambitious endeavor to update all RC records with the full set of RC expression attributes (developmental staging, sexually dimorphic activity, ectopic activity, and enhancer/silencer activity), which did not become a full part of the REDfly data model until the release of REDfly v6 in 2020. Since that time, over one-third of the RC records have been updated to contain this information.

Table 1. REDfly contents as of 1 July 2022.

Reporter Constructs (RCs)	43,819
From in vivo *reporter genes*	*21,690*
Associated with staging and other attributes	*17,961*
CRM_segments	16
Predicted CRMs (pCRMs)	14,318
Inferred CRMs (iCRMs)	7760
Transcription Factor Binding Sites (TFBS)	2717
Publications curated	1366

6. Using REDfly

The extensive data and metadata REDfly provides for each of its records allows for detailed customized searching of the database contents. Typical entry points for a REDfly search are via a gene name (**Figure 2**A(a)) or a literature reference (via PubMed ID, **Figure 2**A(b)). By default, searching for a gene name will execute a “by locus” search in which any elements annotated as being associated with that gene, as well as any elements found within a user-customizable range of 10 kb upstream or downstream of the gene (regardless of assigned target gene), will be returned. Moreover, by default, elements identified solely by assays performed in cultured cells are omitted from the results (**Figure 2**A(d)); unchecking the check-box causes these results to be included.

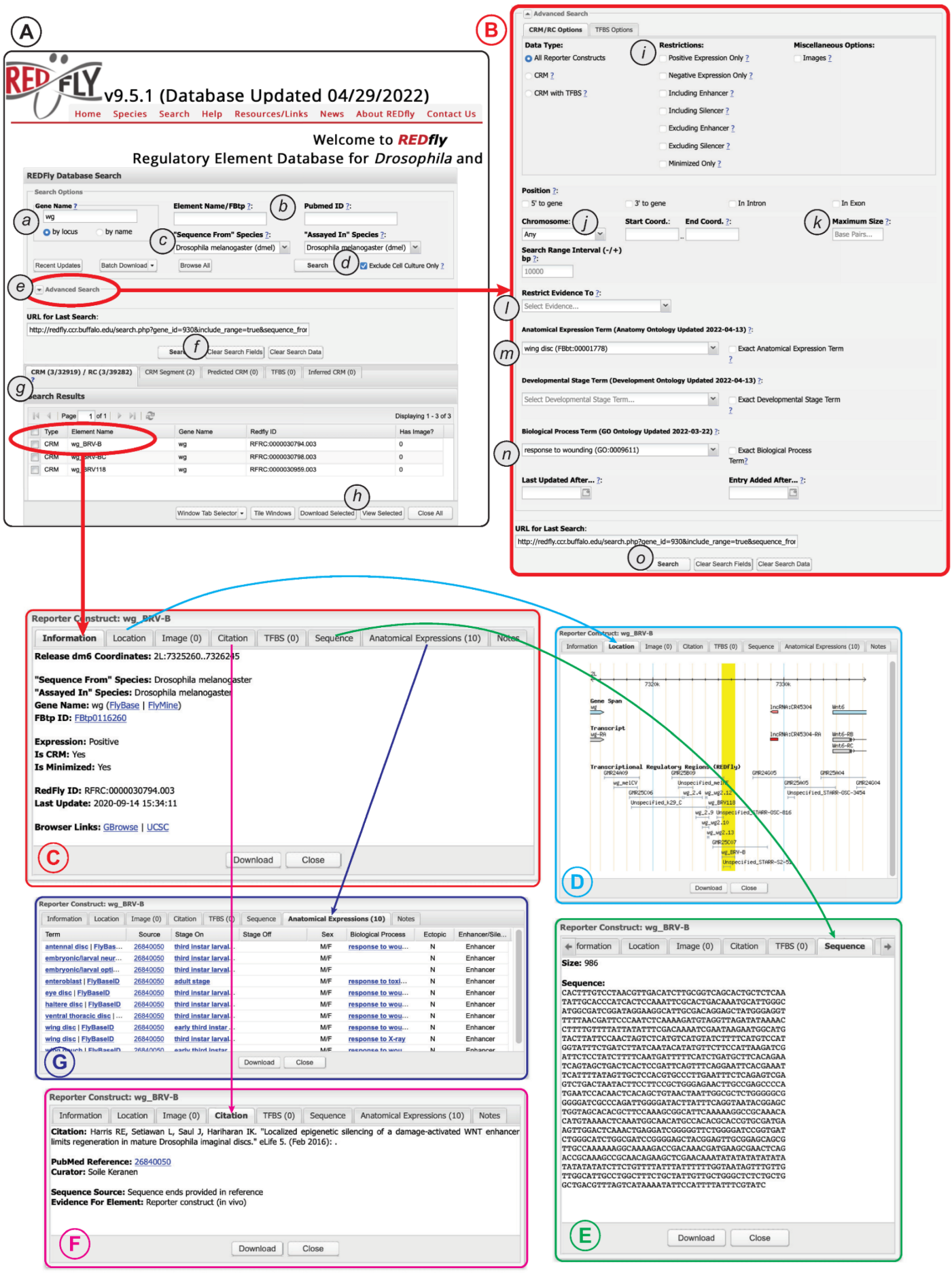
A
v9.5.1 (Database Updated 04/29/2022)
Home Species Search Help Resources/Links News About REDfly Contact Us
Welcome to REDfly
Regulatory Element Database for Drosophila and
REDFly Database Search
Search Options
Gene Name ?
wg
by locus
by name
Element Name/FBtp ?:
Pubmed ID ?:
"Sequence From" Species ?:
Drosophila melanogaster (dmel)
"Assayed In" Species ?:
Drosophila melanogaster (dmel)
Recent Updates
Batch Download
Browse All
Search
Exclude Cell Culture Only ?
Advanced Search
URL for Last Search:
http://redfly.ccr.buffalo.edu/search.php?gene_id=930&include_range=true&sequence_fro
Search
Clear Search Fields
Clear Search Data
CRM (3/32919) / RC (3/39282)
CRM Segment (2)
Predicted CRM (0)
TFBS (0)
Inferred CRM (0)
Search Results
Page 1 of 1
Displaying 1 - 3 of 3
Type
Element Name
Gene Name
Redfly ID
Has Image?
CRM
wg_BRV-B
wg
RFRC:0000030794.003
0
CRM
wg_BRV-BC
wg
RFRC:0000030798.003
0
CRM
wg_BRV118
wg
RFRC:0000030959.003
0
Window Tab Selector
Tile Windows
Download Selected
View Selected
Close All
B
Advanced Search
CRM/RC Options
TFBS Options
Data Type:
All Reporter Constructs
CRM ?
CRM with TFBS ?
Restrictions:
Positive Expression Only ?
Negative Expression Only ?
Including Enhancer ?
Including Silencer ?
Excluding Enhancer ?
Excluding Silencer ?
Minimized Only ?
Miscellaneous Options:
Images ?
Position ?:
5' to gene
3' to gene
In Intron
In Exon
Chromosome:
Any
Start Coord.:
End Coord. ?:
Maximum Size ?:
Base Pairs...
Search Range Interval (-/+) bp ?:
10000
Restrict Evidence To ?:
Select Evidence...
Anatomical Expression Term (Anatomy Ontology Updated 2022-04-13) ?:
wing disc (FBbt:00001778)
Exact Anatomical Expression Term ?
Developmental Stage Term (Development Ontology Updated 2022-04-13) ?:
Select Developmental Stage Term...
Exact Developmental Stage Term ?
Biological Process Term (GO Ontology Updated 2022-03-22) ?:
response to wounding (GO:0009611)
Exact Biological Process Term?
Last Updated After... ?:
Entry Added After... ?:
URL for Last Search:
http://redfly.ccr.buffalo.edu/search.php?gene_id=930&include_range=true&sequence_fro
Search
Clear Search Fields
Clear Search Data
C
Reporter Construct: wg_BRV-B
Information
Location
Image (0)
Citation
TFBS (0)
Sequence
Anatomical Expressions (10)
Notes
Release dm6 Coordinates: 2L:7325260..7326245
"Sequence From" Species: Drosophila melanogaster
"Assayed In" Species: Drosophila melanogaster
Gene Name: wg (FlyBase | FlyMine)
FBtp ID: FBtp0116260
Expression: Positive
Is CRM: Yes
Is Minimized: Yes
RedFly ID: RFRC:0000030794.003
Last Update: 2020-09-14 15:34:11
Browser Links: GBrowse | UCSC
Download
Close
D
Reporter Construct: wg_BRV-B
Information
Location
Image (0)
Citation
TFBS (0)
Sequence
Anatomical Expressions (10)
Notes
Gene Span
Transcript
Transcriptional Regulatory Regions (REDFly)
Download
Close
E
Reporter Construct: wg_BRV-B
formation
Location
Image (0)
Citation
TFBS (0)
Sequence
Size: 986
Sequence:
CACTTTGTCCTAACGTTGACATCTTGCGGTCAGCACTGCTCTCAA
TATTGCACCCATCACTCCAAATTCGCACTGACAAATGCATTGGGC
ATGGCGATCGGATAGGAAGGCATTGCGACAGGAGCTATGGGAGGT
TTTTAACGATTCCCAATCTCAAAAGATGTAGGTTAGATATAAAAC
CTTTTGTTTTATTATATTTCGACAAAATCGAATAAGAATGGCATG
TACTTATTCCAACTAGTCTCATGTCATGTATCTTTTCATGTCCAT
GGTATTTCTGATCTTATCAATACATATGTTCTTCCATTAAGATGG
ATTCTCCTATCTTTTCAATGATTTTCATCTGATGCTTCACAGAA
TCAGTAGCTGACTCACTCCGATTCAGTTTCAGGAATTCACGAAAT
TCATTTTATAGTTGCTCCACGTGCCCTTGAATTTCTCAGAGTCGA
GTCTGACTAATACTTCCTTCCGCTGGGAGAACTTGCCGAGCCCCA
TGAATCCACAACTCACAGCTGTAACTAATTGGCGCTCTGGGGGCG
GGGGATCGCCCAGATTGGGGATACTTATTTCAGGTAATACGGAGC
TGGTAGCACACGCTTCCAAAGCGGCATTCAAAAAGGCCGCAAACA
CATGTAAAACTCAAATGGCAACATGCCACACGCACCGTGCGATGA
AGTTGGACTCAAACTGAGGATCGGGGGTTCTGGGGATCCGGTGAT
CTGGGCATCTGGCGATCCGGGGAGCTACGGAGTTGCGGAGCAGCG
TTGCCAAAAAAGGCAAAAGACCGACAAACGATGAAGCGAACTCAG
ACCGCAAAGCCGCAACAGAAGCTCGAACAAATATATATATATATA
TATATATATCTTCTGTTTTATTTATTTTTTGGTAATAGTTTGTTG
TTGGCATTGCCTGGCTTTCTGCTATTGTTGCTGGGCTCTCTGCTG
GCTGACGTTTAGTCATAAAATATTCCATTTTATTTCGTATC
Download
Close
F
Reporter Construct: wg_BRV-B
Information
Location
Image (0)
Citation
TFBS (0)
Sequence
Anatomical Expressions (10)
Notes
Citation: Harris RE, Setiawan L, Saul J, Hariharan IK. "Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs." eLife 5. (Feb 2016): .
PubMed Reference: 26840050
Curator: Soile Keranen
Sequence Source: Sequence ends provided in reference
Evidence For Element: Reporter construct (in vivo)
Download
Close
G
Reporter Construct: wg_BRV-B
Information
Location
Image (0)
Citation
TFBS (0)
Sequence
Anatomical Expressions (10)
Notes
Term
Source
Stage On
Stage Off
Sex
Biological Process
Ectopic
Enhancer/Sile...
antennal disc | FlyBas...
26840050
third instar larval...
M/F
response to wou...
N
Enhancer
embryonic/larval neur...
26840050
third instar larval...
M/F
N
Enhancer
embryonic/larval opti...
26840050
third instar larval...
M/F
N
Enhancer
enteroblast | FlyBaseID
26840050
adult stage
M/F
response to toxi...
N
Enhancer
eye disc | FlyBaseID
26840050
third instar larval...
M/F
response to wou...
N
Enhancer
haltere disc | FlyBaseID
26840050
third instar larval...
M/F
response to wou...
N
Enhancer
ventral thoracic disc | ...
26840050
third instar larval...
M/F
response to wou...
N
Enhancer
wing disc | FlyBaseID
26840050
early third instar...
M/F
response to wou...
N
Enhancer
wing disc | FlyBaseID
26840050
third instar larval...
M/F
response to X-ray
N
Enhancer
Download
Close

Figure 2. The REDfly search interface. See text for details. (**A**) The basic search panel. (**B**) The Advanced Search panel. (**C**) The Detailed Results “Information” pane. (**D**) The Detailed Results “Location” pane. (**E**) The Detailed Results “Sequence” pane. (**F**) The Detailed Results “Citation” pane. (**G**) The Detailed Results “Anatomical Expression” pane.

Clicking on the “Advanced Search” arrow (**Figure 2**A(e)) allows access to a large variety of additional options (**Figure 2**B), including the ability to restrict searches to specific *RC* attributes, genomic locations, anatomical regions, developmental stages, or biological processes. More detailed and complex search capabilities are under development.

Regardless of whether “basic” or “advanced” search is used, a summary of the results will appear in the “Search Results” pane directly below the main search window (**Figure 2**A(g)). Results for each REDfly data class—*RCs/CRMs*, *CRM_segments*, *pCRMs*, *TFBSs*, and *iCRMs*—are displayed in individual tabs to make it easier to view results by type. Checkboxes allow selection of records for download (**Figure 2**A(h)) in any of a number of convenient formats. Alternatively, clicking on an individual result will open a multipaned “Detailed View” window containing full information for the selected record (**Figure 2**C). Basic location and attribute data are displayed in the “Information” tab, along with links to relevant model-organism databases and genome browsers (**Figure 2**C). The “Location” tab provides a snapshot of the element in its genomic milieu (**Figure 2**D). The “Sequence” tab displays the genomic sequence and its size (**Figure 2**E), while the “Citation” tab (**Figure 2**F) provides a citation and link to the publication describing the current element, plus a description of the evidence used by REDfly curators to annotate sequence and expression information. The “Anatomical Expression” tab (**Figure 2**G) lists each cell type or tissue where the regulatory element is active, along with a specific citation for that activity data (since activity data may be drawn from multiple references), developmental staging for the observed activity, and the other attributes discussed above, e.g., sexually dimorphic activity, ectopic expression, and enhancer or silencer activity.

Since CRM activity can be complex and not easily summarized using the anatomical and staging terms available in the relevant ontologies, the researchers also supply a “Notes” tab containing details and clarifications.

References

1. Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F.; et al. The genome sequence of Drosophila melanogaster. Science 2000, 287, 2185–2195.
2. C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: A platform for investigating biology. Science 1998, 282, 2012–2018.
3. Holt, R.A.; Subramanian, G.M.; Halpern, A.; Sutton, G.G.; Charlab, R.; Nusskern, D.R.; Wincker, P.; Clark, A.G.; Ribeiro, J.M.; Wides, R.; et al. The genome sequence of the malaria mosquito Anopheles gambiae. Science 2002, 298, 129–149.

4. Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002, 420, 520–562.
5. The Alliance of Genome Resources Consortium. The Alliance of Genome Resources: Building a Modern Data Ecosystem for Model Organism Databases. Genetics 2019, 213, 1189–1196.
6. Grosveld, F.; van Staalduinen, J.; Stadhouders, R. Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. Annu. Rev. Genom. Hum. Genet. 2021, 22, 127–146.
7. Chen, D.; Lei, E.P. Function and regulation of chromatin insulators in dynamic genome organization. Curr. Opin. Cell Biol. 2019, 58, 61–68.
8. Segert, J.A.; Gisselbrecht, S.S.; Bulyk, M.L. Transcriptional Silencers: Driving Gene Expression with the Brakes On. Trends Genet. 2021, 37, 514–527.
9. Batut, P.J.; Bing, X.Y.; Sisco, Z.; Raimundo, J.; Levo, M.; Levine, M.S. Genome organization controls transcriptional dynamics during development. Science 2022, 375, 566–570.
10. Kassis, J.A.; Brown, J.L. Polycomb group response elements in Drosophila and vertebrates. Adv. Genet. 2013, 81, 83–118.
11. Atkinson, T.J.; Halfon, M.S. Regulation of Gene Expression in the Genomic Context. Comput. Struct. Biotechnol. J. 2014, 9, e201401001.
12. Abnizova, I.; te Boekhorst, R.; Walter, K.; Gilks, W.R. Some statistical properties of regulatory DNA sequences, and their use in predicting regulatory regions in the Drosophila genome: The fluffy-tail test. BMC Bioinform. 2005, 6, 109.
13. Arnone, M.I.; Davidson, E.H. The hardwiring of development: Organization and function of genomic regulatory systems. Development 1997, 124, 1851–1864.
14. Lifanov, A.P.; Makeev, V.J.; Nazina, A.G.; Papatsenko, D.A. Homotypic regulatory clusters in Drosophila. Genome Res. 2003, 13, 579–588.
15. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 2016, 3, 160018.
16. Li, L.; Zhu, Q.; He, X.; Sinha, S.; Halfon, M.S. Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. Genome Biol. 2007, 8, R101.
17. Papatsenko, D.; Goltsev, Y.; Levine, M. Organization of developmental enhancers in the Drosophila embryo. Nucleic Acids Res. 2009, 37, 5665–5677.

18. Zinzen, R.P.; Girardot, C.; Gagneur, J.; Braun, M.; Furlong, E.E. Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 2009, 462, 65–70.
19. Soluri, I.V.; Zumerling, L.M.; Parra, O.A.P.; Clark, E.G.; Blythe, S.A. Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network. Elife 2020, 9, e53916.
20. Blick, A.J.; Mayer-Hirshfeld, I.; Malibiran, B.R.; Cooper, M.A.; Martino, P.A.; Johnson, J.E.; Bateman, J.R. The Capacity to Act in Trans Varies Among Drosophila Enhancers. Genetics 2016, 203, 203–218.
21. Halfon, M.S. Silencers, Enhancers, and the Multifunctional Regulatory Genome. Trends Genet. 2020, 36, 149–151.
22. Erceg, J.; Pakozdi, T.; Marco-Ferreres, R.; Ghavi-Helm, Y.; Girardot, C.; Bracken, A.P.; Furlong, E.E.M. Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. Genes Dev. 2017, 31, 590–602.
23. Gisselbrecht, S.S.; Palagi, A.; Kurland, J.V.; Rogers, J.M.; Ozadam, H.; Zhan, Y.; Dekker, J.; Bulyk, M.L. Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. Mol. Cell 2020, 77, 324–337.e8.
24. Li, X.Y.; MacArthur, S.; Bourgon, R.; Nix, D.; Pollard, D.A.; Iyer, V.N.; Hechmer, A.; Simirenko, L.; Stapleton, M.; Luengo Hendriks, C.L.; et al. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. 2008, 6, e27.
25. Li, X.Y.; Thomas, S.; Sabo, P.J.; Eisen, M.B.; Stamatoyannopoulos, J.A.; Biggin, M.D. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biol. 2011, 12, R34.
26. Negre, N.; Brown, C.D.; Shah, P.K.; Kheradpour, P.; Morrison, C.A.; Henikoff, J.G.; Feng, X.; Ahmad, K.; Russell, S.; White, R.A.; et al. A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 2010, 6, e1000814.
27. Moshkovich, N.; Nisha, P.; Boyle, P.J.; Thompson, B.A.; Dale, R.K.; Lei, E.P. RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. Genes Dev. 2011, 25, 1686–1701.
28. Bonn, S.; Zinzen, R.P.; Girardot, C.; Gustafson, E.H.; Perez-Gonzalez, A.; Delhomme, N.; Ghavi-Helm, Y.; Wilczynski, B.; Riddell, A.; Furlong, E.E. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. 2012, 44, 148–156.
29. Khoroshko, V.A.; Levitsky, V.G.; Zykova, T.Y.; Antonenko, O.V.; Belyaeva, E.S.; Zhimulev, I.F. Chromatin Heterogeneity and Distribution of Regulatory Elements in the Late-Replicating

Intercalary Heterochromatin Domains of Drosophila melanogaster Chromosomes. PLoS ONE 2016, 11, e0157147.

30. Zhou, J.; Troyanskaya, O.G. Probabilistic modelling of chromatin code landscape reveals functional diversity of enhancer-like chromatin states. Nat. Commun. 2016, 7, 10528.
31. Mateo, L.J.; Murphy, S.E.; Hafner, A.; Cinquini, I.S.; Walker, C.A.; Boettiger, A.N. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature 2019, 568, 49–54.
32. Bozek, M.; Cortini, R.; Storti, A.E.; Unnerstall, U.; Gaul, U.; Gompel, N. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. Genome Res. 2019, 29, 771–783.
33. Ghavi-Helm, Y.; Klein, F.A.; Pakozdi, T.; Ciglar, L.; Noordermeer, D.; Huber, W.; Furlong, E.E.M. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 2014, 512, 96–100.
34. Li, X.; Zhou, B.; Chen, L.; Gou, L.T.; Li, H.R.; Fu, X.D. GRID-seq reveals the global RNA-chromatin interactome. Nat. Biotechnol. 2017, 35, 940–950.
35. Schor, I.E.; Bussotti, G.; Males, M.; Forneris, M.; Viales, R.R.; Enright, A.J.; Furlong, E.E.M. Non-coding RNA Expression, Function, and Variation during Drosophila Embryogenesis. Curr. Biol. 2018, 28, 3547–3561.e9.
36. Mikhaylichenko, O.; Bondarenko, V.; Harnett, D.; Schor, I.E.; Males, M.; Viales, R.R.; Furlong, E.E.M. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. 2018, 32, 42–57.
37. Haines, J.E.; Eisen, M.B. Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo. PLoS Genet. 2018, 14, e1007367.
38. Cusanovich, D.A.; Reddington, J.P.; Garfield, D.A.; Daza, R.M.; Aghamirzaie, D.; Marco-Ferreres, R.; Pliner, H.A.; Christiansen, L.; Qiu, X.J.; Steemers, F.J.; et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 2018, 555, 538–542.
39. Arunachalam, M.; Jayasurya, K.; Tomancak, P.; Ohler, U. An alignment-free method to identify candidate orthologous enhancers in multiple Drosophila genomes. Bioinformatics 2010, 26, 2109–2115.
40. Kantorovitz, M.R.; Kazemian, M.; Kinston, S.; Miranda-Saavedra, D.; Zhu, Q.; Robinson, G.E.; Gottgens, B.; Halfon, M.S.; Sinha, S. Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse. Dev. Cell 2009, 17, 568–579.
41. Kazemian, M.; Zhu, Q.; Halfon, M.S.; Sinha, S. Improved accuracy of supervised CRM discovery with interpolated Markov models and cross-species comparison. Nucleic Acids Res. 2011, 39, 9463–9472.

42. Arbel, H.; Basu, S.; Fisher, W.W.; Hammonds, A.S.; Wan, K.H.; Park, S.; Weiszmann, R.; Booth, B.W.; Keranen, S.V.; Henriquez, C.; et al. Exploiting regulatory heterogeneity to systematically identify enhancers with high accuracy. Proc. Natl. Acad. Sci. USA 2019, 116, 900–908.
43. Aerts, S.; van Helden, J.; Sand, O.; Hassan, B.A. Fine-tuning enhancer models to predict transcriptional targets across multiple genomes. PLoS ONE 2007, 2, e1115.
44. Brody, T.; Yavatkar, A.S.; Kuzin, A.; Kundu, M.; Tyson, L.J.; Ross, J.; Lin, T.Y.; Lee, C.H.; Awasaki, T.; Lee, T.; et al. Use of a Drosophila genome-wide conserved sequence database to identify functionally related cis-regulatory enhancers. Dev. Dyn. 2012, 241, 169–189.
45. Ivan, A.; Halfon, M.S.; Sinha, S. Computational discovery of cis-regulatory modules in Drosophila without prior knowledge of motifs. Genome Biol. 2008, 9, R22.
46. Guo, H.T.; Huo, H.W.; Yu, Q. SMCis: An Effective Algorithm for Discovery of Cis-Regulatory Modules. PLoS ONE 2016, 11, e0162968.
47. Asma, H.; Halfon, M.S. Computational enhancer prediction: Evaluation and improvements. BMC Bioinform. 2019, 20, 174.
48. Su, J.; Teichmann, S.A.; Down, T.A. Assessing computational methods of cis-regulatory module prediction. PLoS Comput. Biol. 2010, 6, e1001020.
49. Kazemian, M.; Suryamohan, K.; Chen, J.Y.; Zhang, Y.; Samee, M.A.; Halfon, M.S.; Sinha, S. Evidence for deep regulatory similarities in early developmental programs across highly diverged insects. Genome Biol. Evol. 2014, 6, 2301–2320.
50. Asma, H.; Halfon, M.S. Annotating the Insect Regulatory Genome. Insects 2021, 12, 591.
51. Clark, A.G.; Eisen, M.B.; Smith, D.R.; Bergman, C.M.; Oliver, B.; Markow, T.A.; Kaufman, T.C.; Kellis, M.; Gelbart, W.; Iyer, V.N.; et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature 2007, 450, 203–218.
52. He, B.Z.; Holloway, A.K.; Maerkl, S.J.; Kreitman, M. Does positive selection drive transcription factor binding site turnover? A test with Drosophila cis-regulatory modules. PLoS Genet. 2011, 7, e1002053.
53. Holloway, A.K.; Begun, D.J.; Siepel, A.; Pollard, K.S. Accelerated sequence divergence of conserved genomic elements in Drosophila melanogaster. Genome Res. 2008, 18, 1592–1601.
54. Macdonald, S.J.; Long, A.D. Fine scale structural variants distinguish the genomes of Drosophila melanogaster and D. pseudoobscura. Genome Biol. 2006, 7, R67.
55. Jiang, P.; Ludwig, M.Z.; Kreitman, M.; Reinitz, J. Natural variation of the expression pattern of the segmentation gene even-skipped in melanogaster. Dev. Biol. 2015, 405, 173–181.

56. Yang, B.; Wittkopp, P.J. Structure of the Transcriptional Regulatory Network Correlates with Regulatory Divergence in Drosophila. Mol. Biol. Evol. 2017, 34, 1352–1362.
57. Khoueiry, P.; Girardot, C.; Ciglar, L.; Peng, P.C.; Gustafson, E.H.; Sinha, S.; Furlong, E.E.M. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. Elife 2017, 6, e28440.
58. Wang, X.F.; Zhou, T.Y.; Wunderlich, Z.; Maurano, M.T.; DePace, A.H.; Nuzhdin, S.V.; Rohs, R. Analysis of Genetic Variation Indicates DNA Shape Involvement in Purifying Selection. Mol. Biol. Evol. 2018, 35, 1958–1967.
59. Peng, P.C.; Khoueiry, P.; Girardot, C.; Reddington, J.P.; Garfield, D.A.; Furlong, E.E.M.; Sinha, S. The Role of Chromatin Accessibility in cis-Regulatory Evolution. Genome Biol. Evol. 2019, 11, 1813–1828.
60. Benton, M.L.; Talipineni, S.C.; Kostka, D.; Capra, J.A. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. BMC Genom. 2019, 20, 511.
61. Halfon, M.S. Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases. Trends Genet. 2019, 35, 93–103.
62. Lindhorst, D.; Halfon, M.S. Reporter gene assays and chromatin-level assays define substantially non-overlapping sets of enhancer sequences. bioRxiv 2022.
63. Kent, W.J. BLAT--the BLAST-like alignment tool. Genome Res. 2002, 12, 656–664.
64. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. Genome Res. 2002, 12, 996–1006.